

Shadow Resistant Direct Image Registration

Daniel Pizarro¹ and Adrien Bartoli²

¹ Department of Electronics, University of Alcalá,
Esc. Politecnica, 28871 Alcalá de Henares, Spain
`pizarro@depeca.uah.es`

² LASMEA, Blaise Pascal University,
24 avenue des Landais, 63411 Aubiere, France
`adrien.bartoli@univ-bpclermont.fr`

Abstract. Direct image registration methods usually treat shadows as outliers. We propose a method which registers images in a 1D shadow invariant space. Shadow invariant image formation is possible by projecting color images, expressed in a log-chromaticity space, onto an ‘intrinsic line’. The slope of the line is a camera dependent parameter, usually obtained in a prior calibration step. In this paper, calibration is avoided by jointly determining the ‘invariant slope’ with the registration parameters. The method deals with images taken by different cameras by using a different slope for each image and compensating for photometric variations. Prior information about the camera is, thus, not required. The method is assessed on synthetic and real data.

Key words: Direct Registration, Shadow Invariant, Photometric Camera Calibration

1 Introduction

The registration of image pairs consists in finding the transformation that best fits two images. That has been a key issue in computer vision, robotics, augmented reality and medical imagery. Although it was thoroughly studied in the past decades, there remain several open problems. Roughly speaking, there are two kinds of approaches: direct and feature based methods. The formers rely on fiducial points described by local properties, which allows matching despite geometric and photometric transformations. The geometric registration is thus formed by minimizing an error between the fiducials position expressed in pixels. As opposed to the local approach, direct methods use pixel discrepancy as a registration error measure. The brightness constancy assumption states that pixel values are equivalent under the sought after transformation. The warp relating two images consists of some geometrical transformation (*e.g.* an homography or an affine transformation) and some photometric model (*e.g.* channel intensity bias and gain or full affine channel mixing).

One of the main problems that arises in direct methods is the existence of partial illumination or shadow changes in the scene to register. In such cases,

the brightness constancy assumption is violated. This paper addresses the problem of directly registering such kind of images. Our proposal is based on expressing the error in a transformed space different from the usual one based on image intensities. In this space which is onedimensional the change of illumination or shadows are removed. This invariant space is governed by a single parameter which is camera-dependent and defines a transformation between the log-chromaticity values of the original RGB image and the invariant image. We propose a method for jointly computing the sought after geometric registration and the parameters defining the shadow invariant space for each image.

Paper Organization

We review previous work and give some background in §2. In §3 we state our error function and give an algorithm for effectively registering images in §4. Results on synthetic and real images are presented in §5. Finally, conclusions are presented in §6.

2 Previous Work

The content of this section is divided into two major parts. First, some previous work about direct image registration is briefly described. The general approach and the most common problems are described. Secondly, some background on color image formation is presented, necessary for describing the process of shadow invariant image formation, which is finally stated.

2.1 Direct Image Registration

The registration of two images is a function \mathcal{P} , which models the transformation between a source image, \mathcal{S} and a target image \mathcal{T} over a region of interest \mathcal{R} . Function $\mathcal{P}(\mathcal{T}(q), q; \phi)$ is parametrized by a vector ϕ composed of geometric and photometric parameters in the general case.

The error function to be minimized make is the sum of square differences of intensity values, over the parameter vector ϕ .

The problem is formally stated as:

$$\min_{\phi} \sum_{q \in \mathcal{R}} \|\mathcal{S}(q) - \mathcal{P}(\mathcal{T}(q), q; \phi)\|^2. \quad (1)$$

A linearization of each residual, which allows to solve it in an iterative Linear Least Square fashion, was popularized by the Lucas-Kanade algorithm [1]. There exist remarkably fast approaches for warps functions forming groups. It is known as the Inverse Compositional algorithm [2] and it has been successfully applied with geometric transformations and in [3] an affine photometric model is also included.

The presence of shadows or illumination changes affect the applicability of equation (1), producing registration errors or divergence in the algorithm. There

exist plenty of proposals in the literature to extend the direct registration with a certain grade of immunity against perturbations. The most common approach is to mark shadow areas as outliers. The use of robust kernels inside the minimization process [4] allows the algorithm to reach a solution. Other approaches try to model the shadows and changes of illumination. In [5] a learning approach is used to tackle illumination changes by using a linear appearance basis.

2.2 Background on Color Image Formation

We present the physical model used to describe the image formation process. The theory of invariant images is described later in terms and under the assumptions stated below.

We consider that all the surfaces are lambertian, that the lights follow a planckian model and that the camera sensor is narrow-band. The RGB color obtained at a pixel is modeled by the following physical model:

$$\rho_k = \sigma S(\lambda_k) E(\lambda_k, T) Q_k \delta(\lambda - \lambda_k) \quad k = 1, 2, 3, \quad (2)$$

where $\sigma S(\lambda_k)$ represents the surface spectral reflectance functions times the lambertian factor. The term $Q_k \delta(\lambda - \lambda_k)$ represents the sensor spectral response function for each color channel k centered at wavelength λ_k . $E(\lambda_k, T)$ is the spectral power distribution of the light in the planckian model. This is modeled by the following expression:

$$E(\lambda, T) = I c_1 \lambda^{(-5)} e^{\left(\frac{-c_2}{T\lambda}\right)} \quad (3)$$

This model holds for a high rank of color temperatures $T = [2500^\circ, 10000^\circ]$. The term I is a global light intensity and the constants c_1 and c_2 are fixed.

According to this model, the value obtained by the camera at any pixel ρ_k is directly obtained by:

$$\rho_k = \sigma I c_1 (\lambda_k)^{-5} e^{\left(\frac{-c_2}{T\lambda_k}\right)} S(\lambda_k) Q_k. \quad (4)$$

2.3 Shadow Invariant Image Theory

The transformation which allows invariant image formation is based on the original work of [6] in which a method for obtaining an illumination invariant, intrinsic image from an input color image is developed. The method relies on the above presented image formation model, based on the assumption of lambertian surfaces, narrow-band sensors and planckian illuminants.

Given the three channel color components $\rho = (\rho_1, \rho_2, \rho_3)$ described in (4), the logarithm of chromaticity ratios are formed.

$$\begin{aligned}\mathcal{X}_1 &= \log\left(\frac{\rho_1}{\rho_3}\right) = \log(s_1/s_3) + (e_1 - e_3)/T \\ \mathcal{X}_2 &= \log\left(\frac{\rho_2}{\rho_3}\right) = \log(s_2/s_3) + (e_2 - e_3)/T,\end{aligned}\tag{5}$$

where $e_k = -c_2/\lambda_k$ only depends on camera spectral response and not on the surface and $s_k = c_1\lambda_k^{(-5)}S(\lambda_k)Q_k$ does not depend on color temperature T .

The pair of values \mathcal{X}_1 and \mathcal{X}_2 lie on a line with direction vector $\bar{e} = (e_1 - e_3, e_2 - e_3)$. Across different illumination temperature T , vector $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)$ moves along the line.

An illumination invariant quantity can be formed by projecting any vector \mathcal{X} onto the orthogonal line defined by $\bar{e}^\perp = (\cos(\theta), \sin(\theta))$. Therefore, two pixels from the same surface viewed under different illuminations get projected at the same place.

To reduce the arbitrary election of the chromaticity ratios, in [6], is proposed a method to use the geometrical mean $(\rho_1\rho_2\rho_3)^{(1/3)}$ of the three channel values as denominator. A vector of three linearly dependent coordinates is obtained. By choosing a proper decomposition, a twodimensional equivalent vector \mathcal{X} is obtained that preserve the essential properties of equation (5).

The transformation \mathcal{L} is simply obtained by projecting vector \mathcal{X} onto the invariant line parametrized by its slope angle θ :

$$\mathcal{L}(\rho, \theta) = \mathcal{X}_1(\rho) \cos(\theta) + \mathcal{X}_2(\rho) \sin(\theta)\tag{6}$$

This transformation, as it has been previously stated, represents the mapping between a color image and its corresponding shadow invariant representation. By explicitly describing the whole color image \mathcal{S} as an input in (6), the result of $\mathcal{L}(\mathcal{S}, \theta)$ is a 1D shadow invariant image. The transformation is therefore global so it does not depend on pixel position $q \in \mathbb{R}^2$, but only on its color value.

The slope angle θ of the invariant line only depends on camera spectral properties, so it varies across different cameras. In [6] it is presented a method to obtain the slope by a calibration step using a color pattern or by a set of preregistered images from the same camera under illumination changes. In [7] an autocalibration approach is presented by finding the slope for which the entropy of the invariant image is minimum. The later method is proved to be capable to find the correct slope with only one image.

The entropy based method unless simple and powerful requires images in which remarkable shadow areas are present. In the case of images in which the change of illumination is global and no shadow is present, the method is not able to produce the correct slope.

3 Joint Image Registration and Photometric Camera Calibration

Image registration is proposed under invariant transformation $\mathcal{L}(\mathcal{S}(q), \theta)$, applied to both the target and the source image.

The new cost function, is expressed as follows:

$$\min_{\phi, \theta_1, \theta_2} \sum_{q \in \mathcal{R}} \|\mathcal{L}(\mathcal{S}(q), \theta_1) - \mathcal{L}(\mathcal{P}(\mathcal{T}, q; \phi), \theta_2)\|^2 \quad (7)$$

As previously stated, the slope parameter θ is, in general, different in both images (θ_1 and θ_2). According to the camera model, any change of illumination intensity and temperature will be discarded in the invariant image once parameter θ is obtained. Therefore, in theory, \mathcal{P} can only be geometrical.

The validity of (7) is based on the assumption that in different cameras, intrinsic images are comparable. However as is stated in this section, in general such an hypothesis does not hold due to differences in camera response functions. A photometric model is proposed for compensating such differences.

3.1 Camera Response Dependent Parameters

In this section it will be shown that besides the slope, between two cameras it is of importance the inclusion of photometric parameters over RGB space so that the invariant space of two images is directly comparable. Such parameters will not try to compensate for global illumination as in previous attempts [3], but instead they will represent a compensation between different camera responses.

Multiple Gain Compensation Assuming that each camera has similar spectral response, so that the values of λ_k are similar, the slope and surface reflectance will produce similar values. However for different channel gains Q_k the log-chromaticity values are affected.

It is thus reasonable to include multiple gains compensation a_k per channel for the target image before computing its log-chromaticity values:

$$\mathcal{X}_k = \log\left(\frac{a_k \rho_k}{a_3 \rho_3}\right) = \log(a_k/a_3) + \log(\rho_k/\rho_3) \quad (8)$$

According to (6), the projection reduces the photometric compensation into a one dimensional offset $d_{\mathcal{L}}$.

$$\mathcal{L}(\rho, \theta) = \log\left(\frac{\rho_1}{\rho_3}\right) \cos(\theta) + \log\left(\frac{\rho_2}{\rho_3}\right) \sin(\theta) + d_{\mathcal{L}}, \quad (9)$$

where $d_{\mathcal{L}} = \log(a_1/a_3) \cos(\theta) + \log(a_2/a_3) \sin(\theta)$.

In the case where both cameras were different by only constant gains, it is still enough as a way to compensate camera responses, to compute a single offset.

Multiple gain and bias for each channel compensation As stated in [8], the real response for most digital cameras is not linear. Under certain range of values we can consider that the camera response can be approximated by a gain and bias function. The presence of bias over RGB represents a problem since the assumption of the invariant line is no longer valid.

Adding bias and gain over RGB results in the following invariant representation:

$$\mathcal{L}(\rho, \theta) = \log\left(\frac{\rho_1 + b_1}{\rho_3 + b_3}\right) \cos(\theta) + \log\left(\frac{\rho_2 + b_2}{\rho_3 + b_3}\right) \sin(\theta) + d_{\mathcal{L}}. \quad (10)$$

Where $d_{\mathcal{L}}$ is the same commented in the multiple gain model, and b_k are biases added to color values.

The total number of required photometric parameters is four under the assumption that only one of the cameras suffer from the bias problem. In the case that both images are suitable to be affected, an extra bias model is introduced for the source image. For such critical case the number of parameters is increased to seven.

The new cost function, which includes photometric parameters (ϕ_p^1, ϕ_p^2) in source and target images, is presented:

$$\min_{\phi, \theta_1, \theta_2, \phi_p^1, \phi_p^2} \sum_{q \in \mathcal{R}} \|\mathcal{L}(\mathcal{S}(q), \theta_1, \phi_p^1) - \mathcal{L}(\mathcal{P}(\mathcal{T}, q; \phi), \theta_2, \phi_p^2)\|^2 \quad (11)$$

Besides the commented models, specially amateur cameras suffer from many artificial perturbations which includes saturation boosting, channel mixing and digital filters applied to the raw image sensed.

4 Minimizing the Error Function

In this section, the optimization process involved in obtaining image registration and invariant space parameters is presented in details.

Given the more general expression (11), which includes a photometric model, a Gauss-Newton approach is derived as an optimization method.

A first order approximation of the warped image around current estimation of parameters $\Phi = (\phi, \theta_1, \theta_2, \phi_S, \phi_T)$ is obtained. The residue in the error function is also renamed by using two different functions \mathcal{W}_1 and \mathcal{W}_2 depending on vector Φ .

The renamed cost function becomes:

$$\min_{\Phi} \sum_{q \in \mathcal{R}} \|\mathcal{W}_1(\mathcal{S}(q), \Phi) - \mathcal{W}_2(\mathcal{T}(q), q, \Phi)\|^2. \quad (12)$$

The Gauss-Newton approximation is:

$$\epsilon^2 \approx \sum_{q \in \mathcal{R}} \|\mathcal{W}_1(\mathcal{S}(q), \Phi) - \mathcal{W}_2(\mathcal{T}(q), q, \Phi) + (L_{\mathcal{W}_1}(q) + L_{\mathcal{W}_2}(q)) \Delta \Phi\|^2, \quad (13)$$

where $L_{\mathcal{W}_1}(q)$ and $L_{\mathcal{W}_2}(q)$ represents respectively the first derivatives of functions \mathcal{W}_1 and \mathcal{W}_2 . ϵ^2 is the residual error from the cost function to minimize.

The parameter increment $\Delta\Phi$ is given by solving the following linear system:

$$E_{\Phi}\Delta\Phi = b_{\Phi}, \quad (14)$$

where E_{Φ} represents the approximated Hessian of the error function:

$$E_{\Phi} = \sum_{q \in \mathcal{R}} (L_{\mathcal{W}_1}(q) + L_{\mathcal{W}_2}(q))(L_{\mathcal{W}_1}(q) + L_{\mathcal{W}_2}(q))^T. \quad (15)$$

The right hand side of the linear system b_{Φ} includes the error image:

$$b_{\Phi} = \sum_{q \in \mathcal{R}} (L_{\mathcal{W}_1}(q) + L_{\mathcal{W}_2}(q))(\mathcal{W}_1(\mathcal{S}(q), \Phi) - \mathcal{W}_2(\mathcal{T}(q), q, \Phi)). \quad (16)$$

Once the increment $\Delta\Phi$ is obtained Φ is updated accordingly with each model.

4.1 Using an Homography for the Geometric Model

The used geometric model consists of an homography transformation. Homographies are fully representative as global geometrical models. They are suitable for registering planar scenes or under camera rotation. It is a groupwise homogeneous transformation represented by a full rank 3×3 matrix H with eight degrees of freedom. The homography is applied to the homogeneous coordinates q in the target image for composing the warp. It is assumed that the first eight coordinates of vector Φ represent the values of ΔH at each iteration.

5 Experimental Results

In this section some of the results are presented in order to validate the proposal. The experiments are designed to compare the convergence properties of our algorithm and to test the photometric models we proposed.

5.1 Synthetic Image Registration

A set of synthetic images is generated according to the model presented in §2. Each image consists of a set of quadrangular color patches under different illuminations, covering a range of color temperatures from 2500° to 10000° . By modifying camera response parameters we are able to simulate images taken by different cameras. Two models are considered for the experiment.

- *Multiple gains*: described by only one bias parameter $d_{\mathcal{L}}$ in the invariant space.
- *Multiple gains and biases*: Considering the complete case, the model has five parameters for the target image $\phi_{p\mathcal{T}}$ and three parameters $\phi_{p\mathcal{S}}$ for the source image.

compared Algorithms In all the experiments the following algorithms are compared:

- DRSI-NP: Direct Registration in Shadow Invariant space with No Photometric model to compensate between cameras.
- DRSI-MG: Direct Registration in Shadow Invariant space with Multiple Gains as a photometric model to compensate between cameras.
- DRSI-MGB: Direct Registration in Shadow Invariant with Multiple Gains and Biases in target image and only bias in source image.
- DR: Direct Registration over greylevel values

Simulation Setup Given two differently illuminated sequences of patches, we simulate a 2D homography by displacing the corners in the target image in random directions by some value γ with default value of 5 pixels. The target image is contaminated by gaussian noise with variance σ and a default value of 25.5. The value of photometric parameters for target and source image has been chosen fixed for the simulations: $\phi_T = (b_1 = 3.2, b_2 = 2.1, b_3 = 1)$ and $\phi_S = (b_1 = 0.8, b_2 = 4, b_3 = 3.5)$. Both target and source image has a slope parameter of $\theta_1 = \theta_2 = 169.23^\circ$. Interest area \mathcal{R} is obtained by using strong edges in greylevel image and dilating them by a factor of 8..

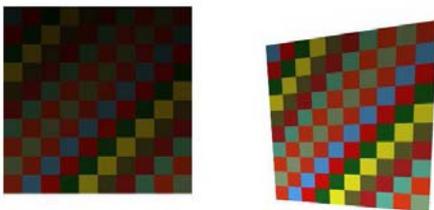


Fig. 1. Pair of synthetic images

Results In Figure 2.a and 2.b the geometric error is presented against noise variance σ and initial pixel displacement γ . In Figure 2.c and 2.d the slope angle error is presented against noise variance σ and initial pixel displacement γ .

5.2 Real Image Registration

For testing the presented proposal with real images, the same planar surface is acquired with two different low-cost commercial cameras. By manually clicking in the four corners of the planar shape, an interest area \mathcal{R} is obtained. If the two regions are far from 10 pixels of displacement a pre-registration is used using manual clicked points. The results presented show the error measured between a pair of images transformed into its respective invariant spaces across the iterations.

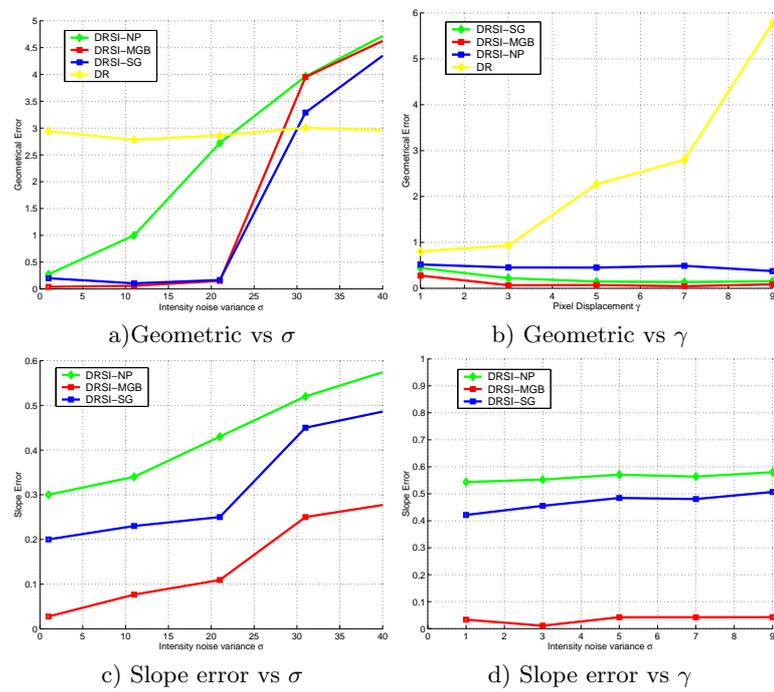


Fig. 2. Residual Error vs σ and Geometrical Error vs γ



Fig. 3. Real images and its resulting registration

6 Conclusions

A new method to achieve direct image registration in the presence of shadows is proposed. The approach is based on minimizing the registration error directly in a transformed space from RGB space. The new space is parametrized by a single camera dependent parameter, the invariant line slope. Such parameter is in general different from each camera, so it is included in the optimization stage. Solving registration parameters in the invariant space from images taken by different cameras offers difficulties due to the response function of each camera. In this paper, two models are proposed to compensate such differences: Multiple Gain compensation and Multiple Gain and Bias. Results on synthetic data show that the last one obtains better registration performance against pixel displacement and noise. In real images, under some conditions the use of the multiple gains and biases model is crucial to achieve registration. If both cameras are of similar response, the simple algorithm which avoid photometric model calculations is the best choice. The use of invariant space in direct methods allows to avoid shadows directly without the need of using complex methods which use illumination modeling or robust kernel optimization.

References

- [1] B. Lucas, T.Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1981.
- [2] S. Baker, I.Matthews. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004.
- [3] A. Bartoli. Groupwise Geometric and Photometric Direct Image Registration. *BMVC'06 Proceedings of the Seventeenth British Machine Vision Conference*, Edinburgh, UK, pp.11–20, September 2006
- [4] S. Baker, R. Gross, I. Matthews, T. Ishikawa, Lucas-Kanade 20 Years ON: A Unifying Framework: Part 2, tech. report CMU-RI-TR-03-01, Robotics Institute, Carnegie Mellon University, February, 2003
- [5] S. Baker, R. Gross, I. Matthews, T. Ishikawa, Face Recognition Across Pose and Illumination, *Handbook of Face Recognition*, Stan Z. Li and Anil K. Jain, ed., Springer-Verlag, June, 2004
- [6] G.D. Finlayson, S.D. Hordley, and M.S. Drew. Removing shadows from images. In *ECCV 2002: European Conference on Computer Vision*, pages 4:823–836, 2002
- [7] G. Finlayson, M. Drew, C. Lu, Intrinsic Images by Entropy Minimization, *Proceedings of 8th European Conference on Computer Vision, Prague, pp 582–595, 2004*
- [8] Kobus Barnard, Brian Funt, Camera characterization for color research, *Color Research and Application*, Vol. 27, No. 3, pp. 153–164, 2002