Experimental Evaluation of Color Illumination Models for Image Matching and Indexing

Patrick GROS
IRISA – CNRS
Campus de Beaulieu
35042 Rennes - France
Patrick.Gros@irisa.fr

Abstract

Recent works have demonstrated that the direct use of grey levels for image matching and indexing allows to build very powerful systems. The present paper tries to enlarge these results to the case of color images.

First it presents a small abstract about photometry and cameras, which allows to justify the choice of a color representation system. Then it presents, evaluates, and compares several illumination models, and discusses image normalization techniques. Finally the paper presents a set of color invariants for image matching and indexing.

The most related works, from which our own work is derived, are Finlayson's and Schmid's. Finlayson [?] introduces illumination invariance as an important criterion for indexing systems. He uses a simple illumination model, without providing a justification, evaluation or comparison with other models. We try to provide such elements in this paper. Schmid [?] uses a set of invariants to match and index grey level images. We propose to adapt this set of invariants to deal with color images.

Our approach is based on both a photometric model of the camera and an experimental evaluation of the proposed models. Thus it should provide stronger arguments about the advantages and drawbacks of the different choices that have to be done to implement a matching or indexing system.

1 Introduction

Content-based image indexing and retrieval was the source of many researches and papers in the last years. Most developed systems are based on a "query by example" paradigm: The user provides an image and the system is supposed to retrieve all the images which are similar in the data base. "Similar" can mean that the images represent the same kind of scene, or it can mean that the same object appears in both images. In the latter case, the background may differ and the shot conditions may vary.

This paper is concerned with the second case of object recognition. Most systems in this domain are based on the search of local descriptors which are invariant to the transformations that are to be discarded in the images [?,?]. The geometric transformations were broadly studied and geometric invariance to rotation, translation, scale and sometimes to affine transformations is achieved in many systems. Even if photometry was already studied by many authors [?,?], especially in the computer graphics or colorimetry communities, many questions are still not solved about its use in image indexing.

One of these questions is the choice of a color information representation system and of an illumination transformation model. The underlying problem is to determine how pixel values change when the illumination conditions vary, and how to get rid of these variations.

In this paper, we propose to revisit this photometric aspect of vision and its application to image matching and image indexing. The light entering a camera is first transformed in a digital signal by CCD receptors and is then subject to a γ correction. We propose to inverse this γ correction and to use a affine model on each of the three resulting linear RGB channels. A set of invariants and a normalization technique are derived from this model and are tested on real color images.

The first section of the paper briefly reminds the bases of photometry. The second one evaluates several illumination models. The third one introduces normalization techniques and the last one local color invariants.

2 A Photometric Model of a Camera

Light is composed of an infinite and bounded set of different wave lengths. The light received by a captor depends on the spectrum emitted by the light sources and the properties of the materials reflecting or transmitting this light. As human eye has only three kinds of color receptors, only three informations are needed to code what a human can visually perceive. Most commercial cameras follow this scheme and have also three kinds of receptors.

The transformation operated by the camera is simple, and involves three parameters. The first one is the reception spectrum of each of the captor, which defines the power received by the captor with respect to the received light. The second one is the response curve with indicates the response of the captor for a given received power. This curve can be approximated by an affine function, bounded by the limits 0 and 255 which cause the saturation problem.

The third parameter is the γ correction [**?**,**?**]. This correction comes from the fact that screens usually use the pixel values to code the voltage input rather than the intensity of the produced light. The relation between these two quantities is non linear:

$$I = (v/v_m)^\gamma h(\lambda)$$

where v_m is the maximum value of the voltage, $h(\lambda)$ is the emitted phosphor spectrum at the maximum voltage and γ is the exponential factor, which is around 2.2 for most color monitors [**?**].

Most cameras precompensate for this non linearity: this process is called γ correction and consists of an approximately 0.45–power function which transform the so-called linear RGB components to the usual non-linear RGB output components. It is easy to get rid of some of the camera parameters: γ correction can be inverted and an affine photometric invariant will compensate for the response curve of the captors. It is not possible to go further without a precise calibration of the camera.

3 Illumination Models

We call illumination model a model which describes how the pixel values of an image vary when the photometric conditions of the shot change. This variation can be three-fold: It can be due to an intensity variation of one of the light source, an variation of the emission spectrum, or a motion of the source.

The first choice to be done is that of a color representation system. Many systems exist [**?**], and are useful for different tasks. RGB provides an information close to the captor technology, but is not adapted to compute color distances. On the other hand, systems like Lab are expensive to compute. Following many authors, we chose to use the RGB system, but in both its linear and non linear flavors.

In this section, we propose to evaluate experimentally several models using both the linear and non linear color representation systems:

M1:	$\mathbf{p}' = \mathbf{p}$	M4:	$\mathbf{p}' = \alpha\mathbf{p} + \mathbf{T}$	M7:	$\mathbf{p}' = \mathbf{M}\mathbf{p}$
M2:	$\mathbf{p}' = \mathbf{p} + \mathbf{T}$	M5:	$\mathbf{p}' = \mathbf{D}\mathbf{p}$	M8:	$\mathbf{p}' = \mathbf{M}\mathbf{p} + \mathbf{T}$
M3:	$\mathbf{p}' = \alpha\mathbf{p}$	M6:	$\mathbf{p}' = \mathbf{D}\mathbf{p} + \mathbf{T}$		

$\mathbf{p} = (r, g, b)^t$ denotes a pixel value which is transformed into $\mathbf{p}' = (r', g', b')^t$, α is a scalar, $\mathbf{D}$ a diagonal matrix, $\mathbf{M}$ a full 3×3 matrix, and $\mathbf{T} = (t_x, t_y, t_z)^t$ a translation vector. Models are denoted with an M (resp. L) when applied to non linear (resp. linear) RGB components. For example, M5 is the model used by Finlayson [?].

In the following tests, we used an image sequence where the camera and the scene do not move. Only the illumination conditions vary from one image to another. A least median square algorithm is used to compute the best parameters which compensate the difference between the first image and the current one. Each model is then evaluated by measuring the remaining differences between both images. To compute this distance, we used L^2. This is not legitimate to measure perceptual differences, but is significant in the case when the corrected image is to be used in further computations (invariants...)

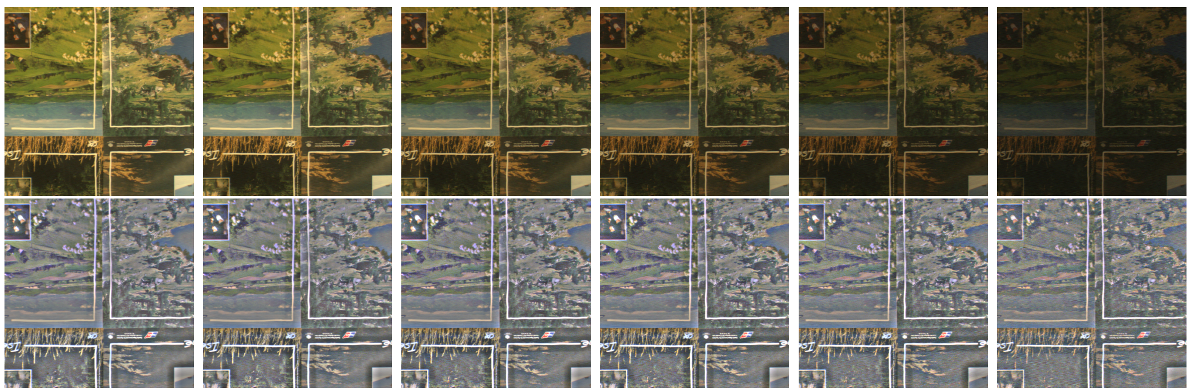

Figure 1: A sequence of images with light intensity variation (up) and the same images after local normalization (bottom).

Model Evaluation under Intensity Variation. In this first test (see table 1), only the intensity of the main light source varies (see Fig. **??**). As this was obtained by using the regulator of an usual halogen lamp, a small variation of the emission spectrum is not impossible. In this case, a model as simple as M3 is enough to correct most of the image differences. More powerful models bring only small improvements. The translation is more important when the image difference is bigger, especially with the L model family.

The difference between L1 and M1 comes from the fact that γ correction modifies the image dynamics. The relative improvement of the result due to L models with respect to the initial difference shown on the L1 line is not bigger than that due to M models with respect to the M1 line. γ correction does not play a big role in this case.

Significance of the Estimated Parameters. The significance of the obtained parameters was tested using a statistical test [?]. The noise affecting the data is assumed to be centered and for each parameter a confidence interval is computed for a confidence level of 95%. The radius of this interval is given by:

$$R(p_i) = \sqrt{\chi^2(95\%, m)}\sqrt{\frac{f}{n}}\sqrt{\sigma_i^2}$$

where $\chi^2(95\%, m)$ is the value of χ^2 distribution for the given confidence level, m is the number of estimated parameters, f is the sum of the errors, n is the number of pixels and σ_i^2 is the variance of the estimated parameter.

If $0 \in [p_i - R(p_i), p_i + R(p_i)]$, 0 appears to be an estimation as good as p_i for the i-th parameter. In such a case the estimation p_i is considered not to be significant.

The significance of model M8 parameters has been tested. Corresponding sub-images of smaller and smaller side (respectively 150, 125, 100, 90, 80, 70, 50, 45, 40, 35, 30, 25, 20, 15, 10, and 5 pixels) were extracted from images 1 and 6. For each pair of corresponding sub-images the model M8 has been estimated, and the significance of each parameter has been computed. The following table shows the minimum sub-image size for which a parameter has been estimated significantly:

parameter	size	parameter	size	parameter	size
$\mathbf{a_{11}}$	5	a_{21}	60	a_{31}	90
a_{12}	60	$\mathbf{a_{22}}$	20	a_{32}	90
a_{13}	125	a_{23}	35	$\mathbf{a_{33}}$	30
$\mathbf{t_r}$	30	$\mathbf{t_g}$	15	$\mathbf{t_b}$	15

The parameters of M6 (in bold in the table) appear to be estimated more significantly even with small sub-images.

Model	Average error between 2 images				
	1 and 2	1 and 3	1 and 4	1 and 5	1 and 6
M1	21.88	64.80	101.63	139.67	180.78
M2	11.25	21.97	33.53	47.70	65.66
M3	10.17	13.61	17.32	21.41	29.45
M4	9.79	11.80	14.59	19.53	29.27
M5	9.97	12.20	15.02	19.45	29.36
M6	9.80	11.58	14.12	18.80	29.01
M7	9.84	11.82	14.39	18.59	28.57
M8	9.74	11.47	13.90	18.36	27.73
L1	17.62	50.14	73.98	95.35	115.80
L2	11.53	29.06	43.79	56.24	69.60
L3	6.99	8.71	10.58	14.08	26.59
L4	6.82	8.33	10.14	13.23	20.01
L5	6.85	7.95	9.59	13.55	25.13
L6	6.84	7.98	9.52	12.46	19.78
L7	6.79	7.88	9.50	13.27	24.32
L8	6.77	7.88	9.37	12.13	18.88

Table 1: Results for an intensity variation.

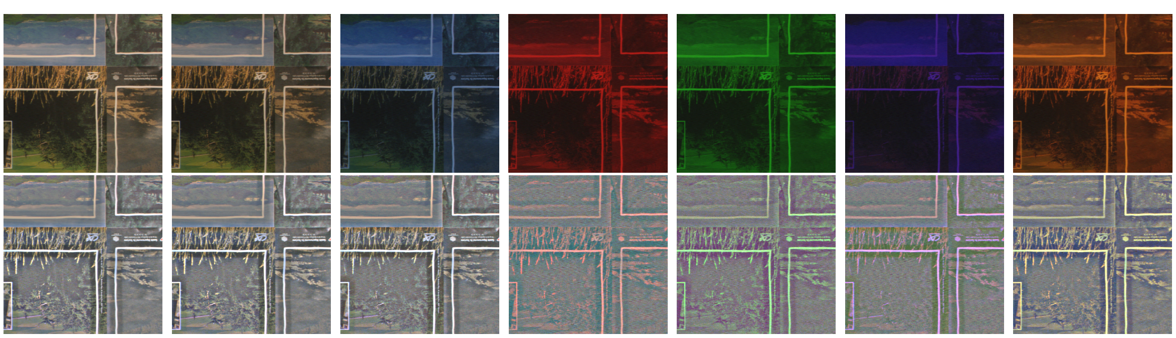

Figure 2: 7 images with variation of the emission spectrum and the same images after local normalization.

Model Evaluation under Spectral Variation. In this sequence (Fig. **??**), we used filters to modify the illumination spectrum of the main source. The images are denoted according to the kind of filter: N = no filter, L = daylight, D = diffused, B = blue, G = green, R = red, Y = yellow (see table 2). It is clear that sophisticated models like M7 or M8 are necessary to obtain a good correction. Inverting the γ correction seems not to improve the results.

Model	Average error between 2 images					
	N / L	N / D	N / B	N / G	N / R	N / Y
M1	90.38	33.31	133.26	136.53	128.04	97.57
M2	35.47	14.98	55.59	63.75	59.82	47.23
M3	65.95	11.03	108.01	109.28	104.30	86.92
M4	31.15	10.71	53.59	60.97	58.25	46.61
M5	17.10	10.92	39.18	53.94	53.83	36.51
M6	17.06	10.70	35.32	51.18	51.87	34.55
M7	16.37	10.70	23.01	25.66	32.43	28.17
M8	16.25	10.65	22.90	25.35	32.15	27.39
L1	53.96	22.56	71.36	71.86	66.11	51.94
L2	34.14	15.00	45.57	50.90	46.29	36.46
L3	45.14	6.37	62.76	56.48	52.72	47.98
L4	31.32	6.39	41.32	45.72	40.57	34.71
L5	11.83	6.36	31.01	41.71	38.51	25.11
L6	10.03	6.34	22.95	35.49	34.78	21.70
L7	11.62	6.29	14.63	18.01	22.12	18.00
L8	9.52	6.37	14.26	17.89	22.03	17.20

Table 2: Results for illumination spectrum variation.

Figure 3: A sequence of images with a moving light source.

Model Evaluation under Light Source Motion. In this case, the test is repeated with a sequence where the light source kept the same spectral emission, but moved around the observed scene (Fig. **??**). The results are not provided because they show that our models are unable to reduce the difference between the images. This is mainly due to the global character of the optimization algorithm used: The illumination variations can have different signs in different parts of the image and a local method is absolutely necessary in this case.

4 Image Normalization

The normalization technique we propose was implemented and tested with the M6 model. It has not been extended to the M7 or M8 models yet. The idea is to get rid of the parameters of the illumination model, which is equivalent to an affine transformation on each of the three RGB channels. This can be achieved by computing the mean and variance of the pixels and to compensate for these parameters to obtain a standard distribution of mean 128 and variance 100 for example.

This process can be applied globally to the whole image. It can also be applied locally: A sub-image of a given size around each pixel is normalized. The normalized value of this central pixel is retained as the value of this pixel in the final image. These two techniques will be denoted $\mathcal{G}$ and $\mathcal{L}$ normalization respectively.

A Few Results. The difficulty of evaluating this technique is that just measuring the error is not enough. The interesting value is the ratio discrimination / invariance, but this ratio is very difficult to evaluate. Here are a few results in terms of remaining differences between normalized images. As these images are not uniformly black, we pretend that they are still discriminant. This is certainly not a formal proof!

The results obtained with the three previous sequences are shown for both $\mathcal{G}$ and $\mathcal{L}$ techniques (see Tab. **??**, **??**, **??**). These results show the interest of a local normalization even in the case where the illumination model is not adapted to the real illumination variation see Tab. **??**).

	Average error between 2 images				
	1 and 2	1 and 3	1 and 4	1 and 5	1 and 6
M1	21.88	64.80	101.63	139.67	180.78
$\mathcal{G}$	17.58	21.00	25.43	33.53	50.77
$\mathcal{L}$	17.90	21.13	24.23	28.76	36.93

Table 3: Difference between two images, and their normalized version using the $\mathcal{G}$ and algorithms. The two original images differ by the intensity of the main light source.

	Average error between 2 images					
	N / L	N / D	N / B	N / G	N / R	N / Y
M1	90.38	33.31	133.26	136.53	128.04	97.57
$\mathcal{G}$	29.97	19.35	65.94	98.55	105.51	66.19
$\mathcal{L}$	25.86	20.38	41.89	50.35	50.19	36.63

Table 4: Difference between two images, and their normalized version using the $\mathcal{G}$ and algorithms. The two original images differ by the emission spectrum of the main light source.

	Average error between 2 images					
	0 and 1	0 and 2	0 and 3	0 and 4	0 and 5	0 and 6
M1	25.16	35.70	56.31	71.62	76.69	81.17
$\mathcal{G}$	53.47	76.54	113.72	138.89	149.08	149.42
$\mathcal{L}$	21.48	23.41	27.40	29.99	31.71	33.18

Table 5: Difference between two images, and their normalized version using the $\mathcal{G}$ and algorithms. The two original images differ by the spatial position of the main light source.

5 Local Color Invariants

In this section invariants using color information are introduced and tested. As far as image parts are concerned, for example to recognize objects, it is necessary to use local tools: Invariants ht

	Seq. 1	Seq. 2	Intensity	W	N
Image points	47.3	260	587.4	99.6	149.3
Matches	28.3	184.4	386.8	56	73,6
Correct matches	28.07	183.3	358.6	40	59
Percentage	99.1%	99.4%	92.7%	71%	80%

Table 6: Matching results.

appear to be a natural choice.

The present work is based on that of Schmid [?]. An improved Harris point detector is used to detect informative points. The signal around each of these points is decomposed on a basis of functions, the derivatives of the Gaussian up to the third order, and mixed in such a way that the obtained descriptors are invariant to rotation and to one of the illumination models. In the presented experiences, we chose the M6 model. This technique provides a set of 23 invariants for each extracted point. These descriptors are either compared to those of another image to match the two images, or stored in a data base to index the image. Translational invariance is obtained by using local coordinates for each detected point; scaling invariance is achieved by a multi-scale approach (see [?]).

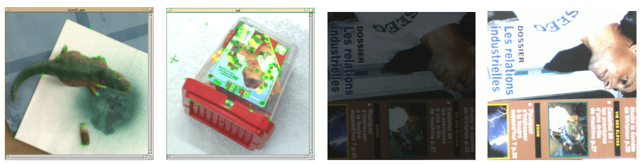

Figure 4: An image of each of the rotational test sequences (left) and 2 images of a sequence with a variation of light intensity (right).

Results with a Scene Rotation. The rotational invariance was tested using two sequences of 30 images taken 6 degrees apart (cf. Fig. ??). In both cases the axis of rotation is the optical axis of the camera. The first image of each sequence was matched with all the other image of the sequence using the rotational invariants. The results are summed up in Tab. ?? (col Seq.1 and Seq.2). Each line provides average numbers for the considered image sequence. The percentage of correct matches is very good. The fact the all points are not matched is due to the poor repeatability of the point detector.

Results for an intensity variation. In this experiment a sequence of 10 images representing a journal cover was used (cf. Fig. ??). The first image was matched to the other ones. The results are presented is Tab. ?? (col Intensity). With intensity variations appears the problem of saturated pixels, which explains a lower rate of correct matches.

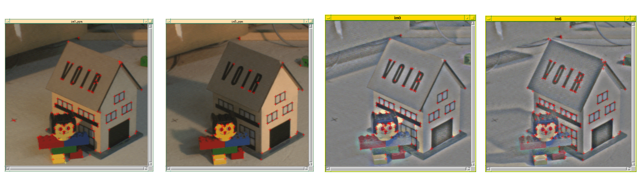

Figure 5: the two original images and the corresponding normalized images.

Results with a moving light source. In this last test, the sequence of Fig. ?? was used.

Between the shots the light source moved around the scene. These images were matched directly (see col. W of Tab. **??**) and using a local normalization technique before the computation of the invariants (see col. N of Tab. **??**). Even if the illumination model used in the normalization technique is not well suited for the case of a moving light source, such a variation is locally equivalent to a simple intensity variation. The effect of normalization is thus twofold: There were more points detected, and these points were detected in a more repeatable way, and the percentage of correct matches is higher (cf Fig. **??**). This example clearly shows the interest of normalization in a practical case.

6 Conclusion

This paper explores the use of color information to match or index images. Various models of color variation under illumination changes are presented and evaluated, several techniques of image normalization are presented, and grey level invariants were extended to the case of color images.

The first results that we obtained are very good and promising, and clearly show the informativeness of the color signal. A particularly interesting possibility is that of finding invariants in the case of a moving light source by erasing most of the shadows. Such a result is unthinkable with grey level images!

References

[FCF96] Finlayson (G.), Chatterjee (S.) et Funt (B.). – Color angular indexing. *In : Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, pp. 16–27.

[FM96] Florou (G.) et Mohr (R.). – What accuracy for 3D measurements with cameras? *In : Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*. pp. 354–358. – IEEE Computer Society Press.

[Lozano98] Lozano (V.). – *Contribution de l'analyse d'image couleur au traitement des images textile*. – Thèse de PhD, Université Jean Monnet, Saint Étienne, January 1998.

[Poynton96] Poynton (C.). – Gamma. *In : A Technical Introduction to Digital Video*. pp. 91–114. – John Wiley & Sons, Ltd.

[Poynton97a] Poynton (C.A.). – Frequently asked questions about color, `http://www.inforamp.net/~poynton/ColorFAQ.html`, 1997.

[Poynton97b] Poynton (C.A.). – Frequently asked questions about gamma, `http://www.inforamp.net/~poynton/GammaFAQ.html`, 1997.

[SM97] Schmid (C.) et Mohr (R.). – Local grayvalue invariants for image retrieval. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, n5, May 1997, pp. 530–534.

[ST97] Sharma (G.) et Trussel (H. Joel). – Digital color imaging. IEEE *Transactions on Image Processing*, vol. 6, n7, July 1997, pp. 901–932.

[TGDK99] Tuytelaars (T.), Gool (L. Van), D'haene (L.) et Koch (R.). – Matching of affinely invariant regions for visual servoing. *In : Proceedings of* IEEE *International Conference on Robotics and Automation*, pp. 1601–1606.