A new cone activation-based natural images dataset

C. Alejandro Parraga, Javier Vazquez-Corral and Maria Vanrell

Computer Vision Center
Dept. de Ciències de la Computació
Universitat Autònoma de Barcelona
Edifíci "O", Campus UAB (Bellaterra)
Barcelona, 08193 Spain

1-Motivation

The analysis of the statistical properties of natural images has yielded a great wealth of knowledge about the operation of the human visual system (HVS). To be useful in many visual science applications, images need to be acquired by a calibrated device, each pixel converted into a device-independent representation (such as the CIE1931 XYZ system) and then converted into cone activation (LMS, for long, medium and short wavelength) representations.

However to produce a truly representative and reliable dataset of cone-activation calibrated natural scenes has proven remarkably difficult.

There are currently several techniques that have produced calibrated databases, which can be summarised as:

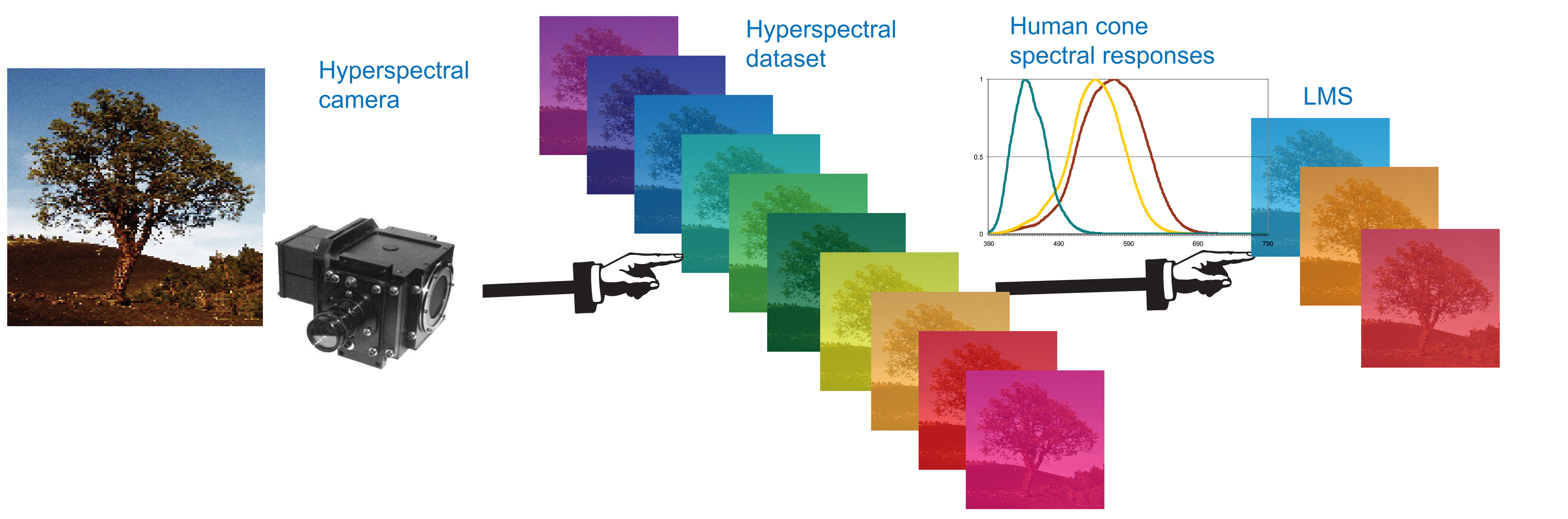
1) *hyperspectral*: exceptionally accurate images, but too few to be called "representative" e.g. publicly available hyperspectral datasets.

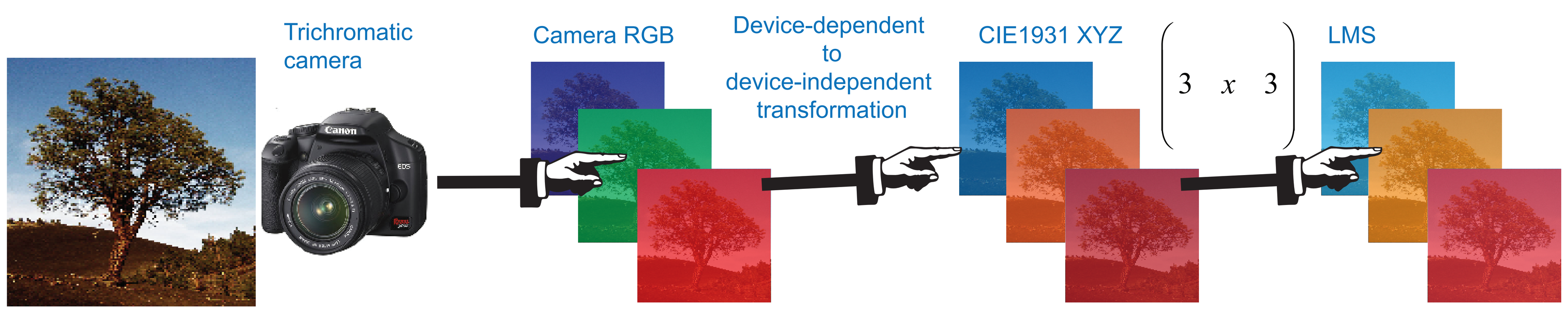
Given the limitations of portability and the need for long integration times and other restrictions of the imaging devices, the first group consists of very limited (typically less than 50) set of samples. Nevertheless, these hyperspectral samples are highly accurate and convey all the necessary information to easily convert every pixel into cone activation representations.

2) *trichromatic*: more representative datasets (i.e. more image samples) based on inaccurate matrix transformations of the camera's sensors into LMS representations.

There several CIE XYZ calibrated databases in the literature, however, in most cases, the spectral sensitivities of the camera sensors are either assumed or estimated, making it very difficult to accurately transform their output into LMS cone representations.

In the cases where the camera sensors are unknown, the conversion between device-dependent RGB representations and LMS representations is done by a simple 3x3 matrix transformation.





2-Methods

We determined the spectral sensitivity of our trichromatic camera (Sigma Foveon SD10) by recording its RGB sensor's responses to light transmitted by a set of 31 spectrally narrowband interference filters. These were later compared to equivalent spectroradiometric measures. The camera's sensors dependency with light intensity and integration time was also measured by means of a Macbeth card.

Other laboratory elements are displayed on the optical bench pictured beside. (a) Constant current power-supply. (B) constant light-source (Tungsten). (c) Reflector box with white target disk. (D) Filter-holder. (E) Camera/spectro-radiometer location. (F) Filters. (G, H) climpex and optical bench apparatus ensuring stability of equipment.

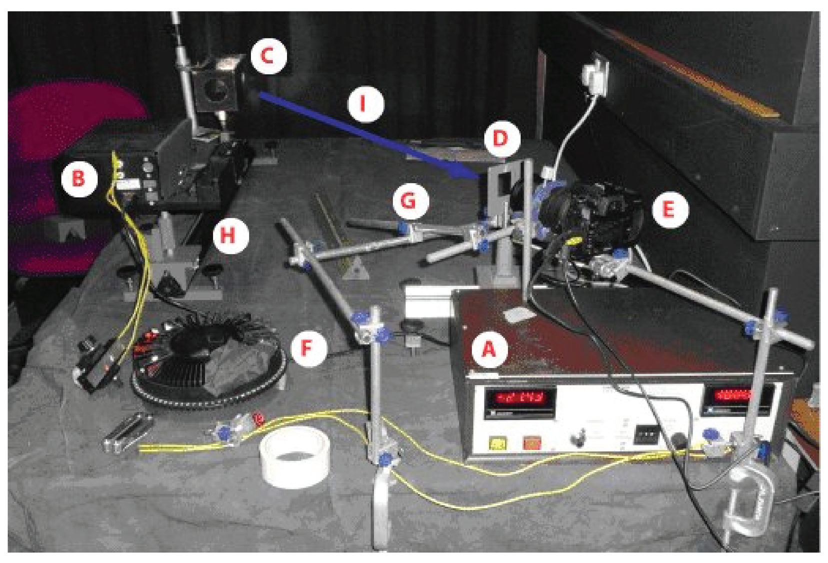
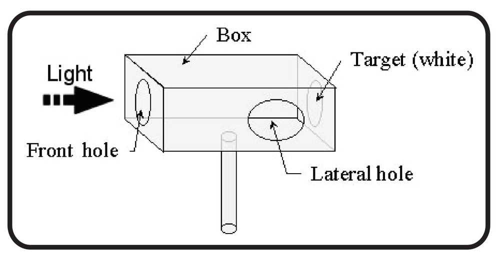
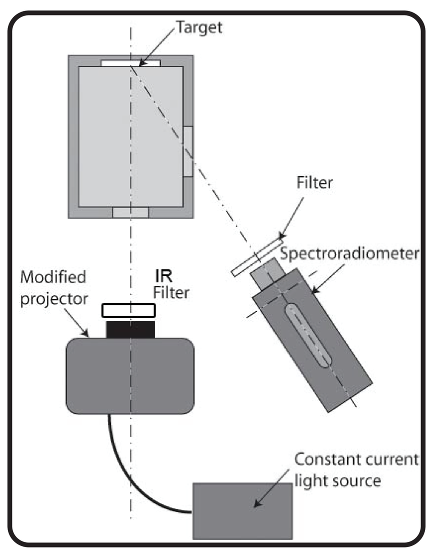
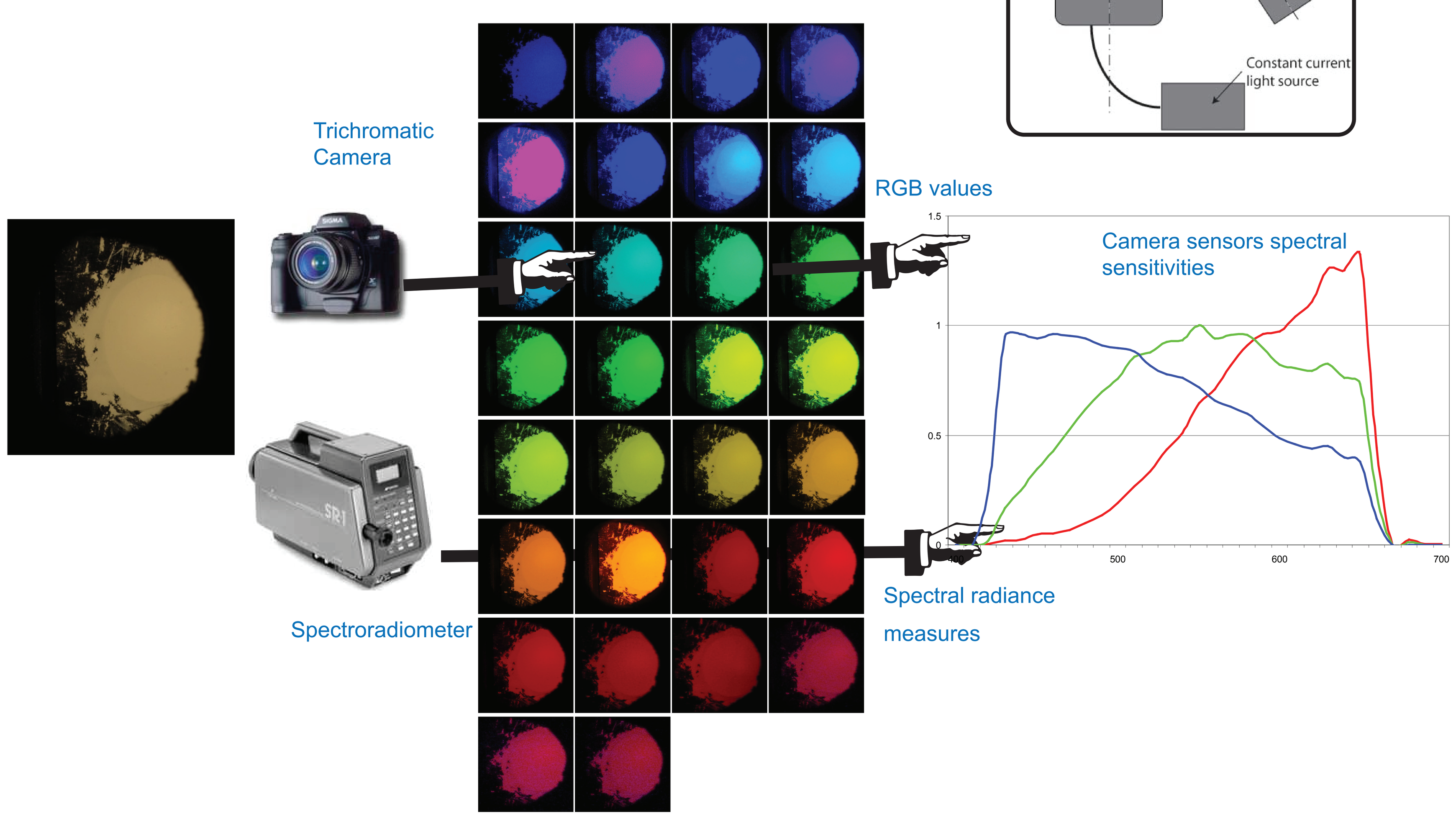


Photo courtesy of George Lovell.





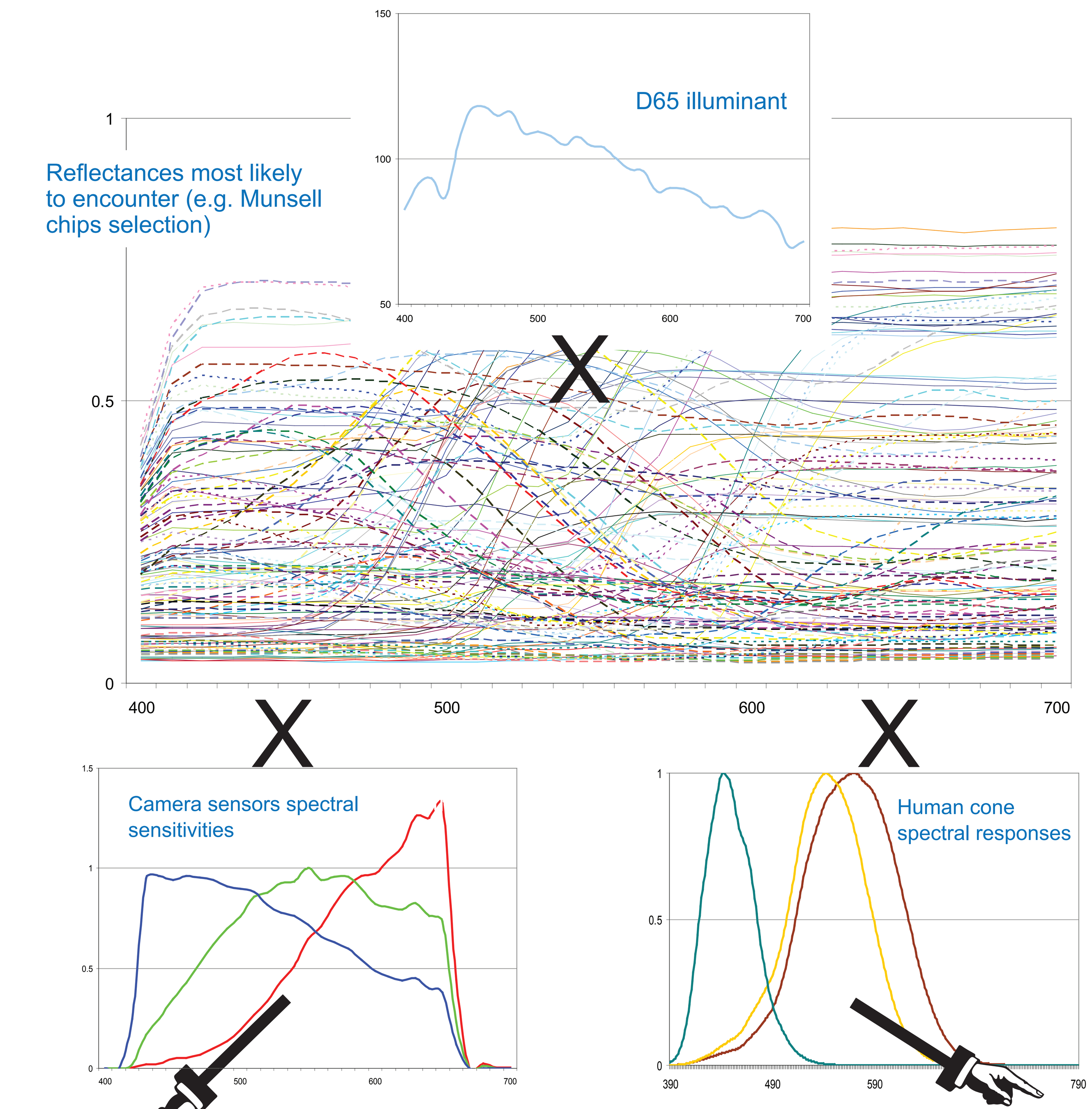


3-Optimisation

The usual transformation from camera colour space to LMS cone activation is done by one of several 3x3 matrix transformations. This transformation is only accurate for monochromatic light (i.e. is only approximate in most visual environments!).

Our solution was to take advantage of the fact that we already know the spectral sensitivity of the camera sensors, which allows us to predict the camera output for a given combination of spectral reflectance and illumination. Once we calculate the sensor's RGB and the corresponding LMS values for our set of spectral radiances of choice, we run a fitting algorithm to find the optimal polynomial transformation between the two of them. The advantage of this solution is that it can be "customised" for any "visual environment" of our choice, be it natural scenes, northern European natural reflectances, man-made pigments, or generalistic environments.

The figure shows an example for a set of 1270 Munsel chip reflectances illuminated by D65 illumination.



$$\begin{vmatrix} R_1 & G_1 & B_1 \\ R_2 & G_2 & B_2 \\ R_3 & G_3 & B_3 \\ \dots & \dots & \dots \\ R_{1270} & G_{1270} & B_{1270} \end{vmatrix}$$

$$\boxed{\begin{array}{l} \varepsilon = \sum_{i=1}^{N} \|J_i - Q_i P\|^2; \\ \text{where } J_i \text{ represents } [L_i, M_i, S_i] \text{ values,} \\ Q_i \text{ is the vector } [R_i, G_i, B_i, R_iG_i, G_iB_i, R_iB_i], \\ \text{and } P \text{ is a 6 x 3 matrix to be determined.} \end{array}}$$

$$\begin{vmatrix} L_1 & M_1 & S_1 \\ L_2 & M_2 & S_2 \\ L_3 & M_3 & S_3 \\ \dots & \dots & \dots \\ L_{1270} & M_{1270} & S_{1270} \end{vmatrix}$$

4-Database

We have generated a dataset of digital natural images where each colour plane corresponds to the human LMS (long, medium, short-wavelength) cone activations. The images were chosen to represent five different visual environments (e.g. forest, seaside, mountain snow, urban, motorways) and were taken under natural illumination at different times of the day. At the bottom-left corner of each picture there was a matte grey ball of approximately constant spectral reflectance (across the camera's response spectrum,) and nearly Lambertian reflective properties, which allows to compute (and remove, if necessary) the illuminant.

The database is also available in CIEXYZ colours

http://www.cat.uab.cat/Datasets/

5-Last note

This new method represents an improvement over the usual 3x3 matrix transformation which is only accurate for spectrally-narrowband colours. The camera-to-LMS transformation can be recalculated to consider other non-human visual systems. The dataset is available to download from our website: http://www.cat.uab.cat/Datasets/

6-Acknowledgements

Supported by projects TIN2007-64577, CSD2007-00018 and RYC-2007-00484 from the Spanish Ministry of Science