Detecting salient cues through illumination-invariant color ratios

Eduardo Todt*, Carme Torras

Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Llorens i Artigas 4-6, 08028 -Barcelona, Spain

Abstract

This work presents a novel technique for embedding color constancy into a saliency-based system for detecting potential landmarks in outdoor environments. Since multiscale color opponencies are among the ingredients determining saliency, the idea is to make such opponencies directly invariant to illumination variations, rather than enforcing the invariance of colors themselves. The new technique is compared against the alternative approach of preprocessing the images with a color constancy procedure before entering the saliency system. The first procedure used in the experimental comparison is the well-known image conversion to chromaticity space, and the second one is based on successive lighting intensity and illuminant color normalizations. The proposed technique offers significant advantages over the preceding two ones since, at a lower computational cost, it exhibits higher stability in front of illumination variations and even of slight viewpoint changes, resulting in a better correspondence of visual saliency to potential landmark elements.

Keywords: visual landmarks, color constancy, visual saliency, visual robot navigation

* Corresponding author
E-mail address: todt@ieee.org

1 Introduction

The extraction of reliable visual landmarks for robot localization in outdoor unstructured environments is still an open research problem. One of the main difficulties is that acquired visual information is strongly dependent on *lighting geometry* (direction and intensity of light source) and *illuminant color* (spectral power distribution), which change with sun position and atmospheric conditions [29]. In order to overcome these adversities, the acquired images are often submitted to transformations, in an attempt to reduce the dependence on illumination. This desired invariance of color representation to general changes in illumination is called *color constancy* [1, 6, 21].

In this work, we evaluate three approaches to color constancy applied to a visual landmark detection system. The first two approaches use standard color constancy preprocessing algorithms followed by the landmark detection, while for the third approach, which is the main contribution of this work, we designed a novel color constancy algorithm embedded in the landmark detector.

The paper is organized as follows. In Section 2, the visual saliency and opponent color concepts are introduced, followed by a description of the landmark detection system based on visual saliency. The adopted color model and color constancy techniques used as preprocessing stages are explained in Section 3, together with their connection to the landmark detection system. In Section 4, the proposed visual saliency algorithm, enhanced with embedded color constancy based on color ratios, is described. Finally, in Section 5, all techniques are discussed and compared in the context of saliency-based landmark detection.

2 Saliency-based landmark detection

When there is no exact knowledge of what things in the environment can be used as landmarks for visual robot localization, some criteria are needed to decide which regions in the images can potentially represent good landmarks. Our proposal is to apply a biologically-inspired visual saliency mechanism to detect potential landmark locations in acquired images.

This section describes the concept of visual saliency and the system we will use to compute visual saliency based on opponent colors.

2.1 Visual saliency

Human vision and artificial vision have in common the challenge of reducing the amount of sensorial information to be processed in order to analyze a scene image, due to limitations in bandwidth, memory, and computational speed. The most accepted models of the human visual system [11, 27] consider the existence of an attention mechanism responsible for selecting the most relevant visual stimuli for further processing by the available resources, rather than attempting to fully interpret visual scenes in a parallel fashion. The attention mechanism is driven by the *visual saliency* of the scene elements, which refers to the idea that certain parts of a scene are distinctive and that they create some form of significant visual arousal at the early visual stages [15]. This mechanism is essentially data-driven, which is particularly useful in those situations where the semantics of the contents of the image is not known and a model of the perceived objects is not available [22].

Light intensity contrast appears to be the primary variable on which humans base visual saliency computation. At higher processing levels in the visual cortex, other feature dimensions participate in defining visual saliency. Among these are edge or line

orientation, color, motion, and stereo disparity. One major observation is that in each case the relevant variable is not the amplitude of visual signals in a particular feature dimension, but the contrast between this amplitude at a given point and at the surrounding locations [28].

Therefore, the notion of saliency relies on the previous notion of opponency. A red roof is salient in a green landscape, but not if it is surrounded by similarly reddish walls and terraces. Likewise, a vertical pole is salient if it is in the middle of a horizontally stripped fence.

Thus, we adhere to the following definition of saliency: given pairs of opponent features (to be introduced in the sequel), a region in an image is considered *salient* if it ranks high in a given feature and its surround ranks high in the opposite feature.

An important characteristic is that saliency is not necessarily associated with a specific feature. For example, a red line among green lines can be as salient as a vertical line among horizontal ones. This allows quantification of saliency measures from different features and their comparison with respect to one another [19].

For each opponent feature pair, saliency is computed by center-surround difference operations, reproducing the model of visual receptive fields [12]. To compute the differences, the Enroth-Cugell and Robson's model [3] is adopted, which considers the effect of the light weighted according to the distance to the center of the receptive field by a difference of Gaussian functions.

2.2 Opponent colors

In the late 19th century, the German physiologist Ewald Hering laid the foundations of color opponency theory, which sustains the existence of three opponent processes in the human visual system, constituted of red-green, yellow-blue and intensity (black-

white) channels [21]. The opponent-color components $R_oG_oB_oY_o$ are calculated from the input *RGB* as follows, taking only positive values [12]:

$$R_o = R - (G + B)/2 \quad (1)$$

$$G_o = G - (R + B)/2 \quad (2)$$

$$B_o = B - (R + G)/2 \quad (3)$$

$$Y_o = (R + G)/2 - |R - G| - B \quad (4)$$

The resulting opponent color image is then processed with the visual saliency system described in Section 2.3. Other definitions of opponent colors have been proposed in the literature, as for example, disregarding the term |*R*-*G*| in the computation of the yellow [24, 30], using logarithmic differences and color ratios [2, 9], or minimizing the correlation between the color components [17, 20, 25]. We tested all these definitions, and found the adopted formulation (1)-(4) better than the others for our system.

2.3 A multiresolution saliency-based system for detecting potential landmarks

Figure 2 shows a diagram representing our complete visual saliency detection system. There are three parallel vertical data flows, each corresponding to a feature type considered, namely intensity, orientations and color opponencies. The input *RGB* image is optionally submitted to a color constancy preprocessing, and subsequently the opponent colors are extracted.

Gaussian pyramids [4] corresponding to the color components are constructed, eight spatial resolutions being represented in each of them. In these pyramids, each level is obtained by a low-pass filtering operation on the preceding level, followed by a subsampling of factor two in each dimension. Level 0 corresponds to the finest scale image and the level 7 to the coarsest image.

The low-pass filtering is computed using a separable cubic B-spline mask with five elements [1, 4, 6, 4, 1], which provides a good Gaussian approximation with low computational cost [13].

In these pyramids, due to the successive low-pass filtering and subsampling operations, a pixel at a fine scale *c* corresponds to a center region, whereas the respective pixel at a coarser scale *s* corresponds to its surround. Then, the center-surround differences, denoted by Θ, can be computed by interpolation to the finer scale and single differences between corresponding pixels at fine and coarse scales within the pyramids.

Center-surround differences are determined for all features at different scale combinations, resulting in partial visual saliency maps. Using several scales, not only for *c* but also for *s*, yields truly multiscale feature extraction, it being possible to detect visual salient objects within a wide size range. The resultant partial maps are combined into a global map, in which salient areas are indicated by large values, whereas non-salient areas have small values.

2.3.1 Partial saliency maps

For the intensity feature *I*, a set of partial saliency maps $SM_I(c,s)$ is constructed detecting either dark centers on bright surrounds or bright centers on dark surrounds, using as centers pixels at pyramid levels $c \in \{2,3,4\}$ and, as their corresponding surrounds, pixels at levels $s = c + d$, $d \in \{3,4\}$:

$$SM_I(c,s) = \left| I(c) \ \Theta \ I(s) \right| \tag{5}$$

For the color opponency features, a set of partial saliency maps is constructed with a double opponency mechanism: in the center regions one color component (e.g., red) contributes to increase the saliency and its opponent color (e.g., green) inhibits the

saliency, while the converse is true in the surround region. Such saliency is defined for the red/green, green/red, blue/yellow, and yellow/blue color pairs, using as centers pixels at pyramid levels $c \in \{2,3,4\}$ and as their corresponding surrounds pixels at levels $s = c + d$, $d \in \{3,4\}$, as follows:

$$SM_{RG}(c,s) = \left| \left(R(c) - G(c)\right) \Theta \left(R(s) - G(s)\right) \right| \quad (6)$$

$$SM_{BY}(c,s) = \left| \left(B(c) - Y(c)\right) \Theta \left(B(s) - Y(s)\right) \right| \quad (7)$$

For the orientation feature, a set of partial saliency maps that represents the local orientation contrast between center and surround scales is built as follows, using as centers pixels at pyramid levels $c \in \{2,3,4\}$ and as their corresponding surrounds pixels at levels $s = c + d$, $d \in \{2\}$:

$$SM_{O}(c,s,\theta) = \left| O(c,\theta) \ \Theta \ O(s,\theta) \right| \quad (8)$$

In total, using the specified *c* and *s* center-surround scales, and the four orientations, 30 partial saliency maps are built, six for intensity, 12 for color, and 12 for orientation. These partial saliency maps need to be combined to obtain one global saliency map. They cannot be simply added, because one salient region present in only a few maps can be masked by noise or less salient regions present in a larger number of maps.

The overall saliency in the scope of individual partial saliency maps is also important. A map with a small number of strong saliency peaks has more relevance than another map with a large number of comparable saliency peaks (Figure 1).

Considering the issues above, the process of combining the partial saliency maps was structured in two stages, *map normalization* and *map weighting*, as described in the following.

2.3.2 Map normalization

In this stage, the partial saliency maps of each feature type (color, intensity and orientation) are normalized by the maximum saliency value obtained at all center-surround scales for the corresponding feature. This transformation equalizes the saliency of different feature types, preserving their relative saliency among scales, and it is computed as follows:

$$\overline{SM}_f(c,s) = \left(\frac{SM_f(c,s)}{\max\left(SM_f(c_i,s_j)\right)} \right) * 255 , \quad (c_i,s_j) \in C_f \times S_f \tag{9}$$

where the subscript *f* is the feature type (color, intensity and orientation), $\overline{SM}_f(c,s)$ is the normalized saliency map for feature *f* for center-surround scales *c* and *s*, $SM_f(c_i,s_j)$ is the saliency map for feature *f* for center-surround scales c_i and s_j, and C_f and S_f are the sets of center and surround scales computed for feature *f*. The constant 255 was introduced only for compatibility with the standard eight-bit-per-pixel gray-level images usually found in image processing applications, and it has no influence on the final results.

2.3.3 Map weighting

In a second stage, the maps are weighted by their information content, taking into account their ability to discriminate the salient regions.

One well-known approach to determine the information content of an image is based on the zeroth order entropy measure [14]. According to Shannon's definition of entropy [23], given a vector *v* of elements from a discrete random variable with n possible classes {*x1, x2, ..., xn*}, where the probability that $x_i \in v$ is $p_i = P(x_i)$, the entropy *H* of *v* if given by:

$$H(v) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{10}$$

The image, here the partial saliency map, corresponds to the vector *v*, with the gray level value of each pixel being the value of the discrete variable, and the probability of each possible value of the variable is approximated by the normalized histogram of the image. The number of bins in the histogram is fixed to 256, corresponding to the number of discrete gray levels in the normalized images. This selection is not critical for our application, and 128 and 64 bins have also been tested, resulting in similar results.

The entropy of an image is maximum for a uniform distribution, and it corresponds to the number of bits needed to represent all pixel values. For example, a uniform distribution consisting of 256 gray levels has maximum entropy $H_{\max} = -\sum_{i=1}^{256}(1/256) * \log_2(1/256) = 8$. In the other extreme, the entropy of an image is minimum for a distribution concentrated in just one value, and its value is zero.

Thus, saliency maps having lower entropy have the saliency values more unevenly distributed than saliency maps with higher entropy (Figure 1). Since the conspicuity of the saliency regions has inverse relation to the uniformity of saliency, we propose to weight the saliency maps considering the maximum entropy of the set of saliency maps and the entropy of each map, according to the formula:

$$W_h(SM_i) = (W_{h\max} - 1) * \left[1 - H(SM_i) / \max H\right] + 1 \tag{11}$$

where $W_h(SM_i)$ is the entropy weight of the *ith* partial saliency map, $W_{h\max}$ is a constant specifying the maximum value for the entropy weight, and $\max H$ is the maximum entropy of the saliency map set considered. The resultant weight is restricted to the range $\left[1, W_{h\max}\right]$.

Although the normalization of the partial saliency maps was considered necessary to combine saliency maps derived from different features, we observed that it was desirable to preserve some amount of the strength of the detected saliencies. For example, if some red object is very salient in a *RG* saliency map, and another object, blue, is also salient in a *BY* map, but not so salient as the former, it is interesting to preserve this relation of saliency strengths, in order to indicate that the red object is more salient than the blue object. Clearly there is a trade-off between the normalization process and the meaning of the absolute saliency values, and it is necessary to introduce some mechanism to deal with this issue.

The solution proposed in this work to solve this trade-off is to allow a modulation of the normalized saliencies by the maximum value of saliency present on the respective partial saliency maps. Thus, the partial saliency maps are also weighted according to the following:

$$W_s(SM_i) = (W_{s\max} - 1) * \left[\max(SM_i) / \max S\right] + 1 \tag{12}$$

where $W_s(SM_i)$ is the saliency weight of the i^{th} partial saliency map, $W_{s\max}$ is a constant specifying the maximum value for the saliency weight, and $\max S$ is the maximum saliency value present on the partial saliency maps considered, before normalization. The resultant weight is restricted to the range $\left[1, W_{s\max}\right]$.

The weight of each normalized partial map is determined by the product of the entropy and saliency weights:

$$W(SM_i) = W_h(SM_i) * W_s(SM_i) \tag{13}$$

2.3.4 The global saliency map

Finally, the normalized partial saliency maps are subject to exponentiation, weighted with $W(SM_i)$, and added to compose the global saliency map:

$$SM = \sum_i W(SM_i) * e^{SM_i} \quad (14)$$

The most salient image regions in this map are subsequently analyzed to either discard them as useful landmarks or to obtain visual signatures, capable of identifying each of them as an existing or a new landmark. A detailed description of landmark characterization and retrieval is beyond the scope of this paper, and we just give a brief account of them in the remaining of this paragraph. For each salient region, three concentric spatial regions are defined: (1) the saliency spot, obtained using local-maxima segmentation of the saliency map, (2) the adjusted landmark region, obtained with backprojection and mean-shift of the saliency map and the chromaticity image, and (3) the surround region, obtained through expansion of the adjusted landmark regions. Descriptors of these regions are computed, based on color and saliency histograms, and they are compared with other landmark descriptors using quadratic-form distance metrics. Descriptors with a low distance to an already stored descriptor are considered different acquisitions of the same landmark, while descriptors without a matching peer are considered new landmarks. Descriptors with poor color information are discarded, because the retrieval system is based on color and saliency distributions.

Considering that the goal of this work is to compare different color constancy techniques applied to saliency detection, visual saliency is computed in what follows based only on color information, disregarding intensity and orientations, although these also play an important role in the complete visual saliency system [26].

3 Color constancy as a preprocessing stage

This section describes the color model adopted in this work, together with two approaches to make the visual saliency system more robust to illumination changes using color-constancy preprocesssing.

3.1 Color model adopted

The color analysis performed in this work is based on the physics-based *dichromatic reflection model* [16], which describes the light reflected from an infinitesimal surface patch of an inhomogeneous dielectric object as a linear combination of light from specular reflection (surface reflection) and diffuse reflection (body reflection). The light reflected on the surface has approximately the same spectral power distribution as the light source. The light that is not reflected at the surface penetrates into the material body, where it is scattered and selectively absorbed. Some fraction of this light arrives again at the surface and exits the material. This body reflection represents the characteristic object color. According to [16],

$$C = m_b(\vec{n},\vec{s})\int_\lambda f_c(\lambda)e(\lambda)c_b(\lambda)d\lambda + m_s(\vec{n},\vec{s},\vec{v})\int_\lambda f_c(\lambda)e(\lambda)c_s(\lambda)d\lambda \qquad (15)$$

where C is the light sensor response corresponding to a surface patch illuminated by an incident light $e(\lambda)$, λ is the light wavelength, m_b and m_s are the geometric dependencies on body and surface, $\vec{n}$ is the surface normal, $\vec{s}$ is the direction of illumination source, $\vec{v}$ is the direction of the viewer, $c_b(\lambda)$ and $c_s(\lambda)$ are the body and surface spectral reflection properties, and $f_c(\lambda)$ represents the spectral sensitivity of the sensor c. For acquiring color images, usually three sensors are used, with their maximum sensitivity in the red, green and blue parts of the visible spectrum. The integration of the light information for each sensor results in a three-dimensional vector [R, G, B].

The angular distribution of the surface reflected light component tends to be strongly peaked around the specular direction, causing highlights of surface reflected light. In general, the surface reflected highlights are localized both in position and direction, resulting in a dominance of body reflection. Since outside the specular peaks the body

reflected light dominates the scene radiance, here it is possible to use a simplified *unichromatic reflection model*, with only the body reflection component represented:

$$C = m_b(\vec{n},\vec{s})\int_\lambda f_c(\lambda)e(\lambda)c_b(\lambda)d\lambda \tag{16}$$

Assuming narrow-band sensors, whose spectral responses can be approximated by delta functions $f_c(\lambda) = \delta(\lambda - \lambda_c)$, the measured sensor responses are:

$$C = m_b(\vec{n},\vec{s})e(\lambda_c)c_b(\lambda_c) \tag{17}$$

This narrow-band sensor assumption is present in several works related to color processing [2, 8, 10, 18] because, being a reasonable approximation, it simplifies a lot the reasoning about color constancy. Finlayson, Drew and Funt [5] proposed to use a linear combination of sensor sensitivities to obtain virtual sensors with sharper responses, thus reducing the error due to the narrow-band assumption.

Unless explicitly noted, in the following discussion the unichromatic reflection model and a camera with three narrow-band sensors *RGB* are assumed.

3.2 Using lighting intensity normalization as a preprocessing stage for visual saliency

The first color constancy technique used at a preprocessing stage that we consider is the transformation of the *RGB* space to chromaticity coordinates (*rgb*) [29]:

$$r = R/(R+G+B) \tag{18}$$

$$g = G/(R+G+B) \tag{19}$$

$$b = B/(R+G+B) \tag{20}$$

Substituting equation (16) in the *r* expression above,

$$r = \frac{m_b(\vec{n},\vec{s})\int_\lambda f_R(\lambda)e(\lambda)c_b(\lambda)d\lambda}{m_b(\vec{n},\vec{s})\left[\int_\lambda f_R(\lambda)e(\lambda)c_b(\lambda)d\lambda + \int_\lambda f_G(\lambda)e(\lambda)c_b(\lambda)d\lambda + \int_\lambda f_B(\lambda)e(\lambda)c_b(\lambda)d\lambda\right]} \tag{21}$$

If we assume a white light source, $e(\lambda)$ is constant over all frequencies. Then, the integration of each sensor response f_c and body reflectance c_b also gives constant values, denoted k_R, k_G, and k_B. In this context, these constants correspond to the scalar responses for the red, green and blue sensors. The dependencies on illumination, surface normal and illumination direction are factored out, resulting in an expression only dependent on the sensor spectral characteristics and the body reflectance:

$$r = k_R /(k_R + k_G + k_B) \tag{22}$$

The same substitution can be applied to the *g* and *b* coordinates:

$$g = k_G /(k_R + k_G + k_B) \tag{23}$$

$$b = k_B /(k_R + k_G + k_B) \tag{24}$$

The pixels with very low intensity provide unstable chromaticity information. Therefore, a common practice is to mask low-intensity pixels when applying the chromaticity transformation. Some authors use a threshold of 1/10 of the maximum image value [12] and others apply an absolute threshold of 30 to the sum of *RGB* values [20]. In our implementation, the pixels with intensity lower than 1/10 of the maximum intensity are assigned a zero *rgb* value. These pixels define a mask that is used to build a masking Gaussian pyramid, where each level is used to mask the partial saliencies computed at the corresponding center-surround scale combination. With this scheme, false saliencies produced by regions with low intensities are avoided. The masking pyramid is an improvement over a simple threshold, because it avoids false saliencies between regions and their surrounds at multiple scales.

The colors represented in *rgb* coordinates are much more stable to lighting changes than those in the *RGB* space, because the light intensity component is removed from each pixel. However, they fail to be invariant under spectral power distribution changes

of the light source, because this type of perturbation affects the response of the *RGB* sensors in different proportions.

3.3 Lighting intensity and illuminant color normalizations as a preprocessing stage for visual saliency

In order to overcome the unfavorable sensitivity to changes in illuminant color shown by the previous normalization, Finlayson, Schiele, and Crowley [8] proposed an algorithm for color constancy called *comprehensive color normalization*, based on iterating two types of successive color normalizations. These normalizations are aimed at removing dependence on both lighting intensity and illuminant color, in an alternate manner.

The first normalization is the same as before, transforming the image to chromaticity coordinates. The second normalization transforms each pixel according to the global mean value of the color bands:

$$r' = r/(3*\overline{r}) \tag{25}$$

$$g' = g/(3*\overline{g}) \tag{26}$$

$$b' = b/(3*\overline{b}) \tag{27}$$

where $\overline{r}$, $\overline{g}$ and $\overline{b}$ are the mean value of the red, green and blue bands in the whole image. The effect of the second normalization can be verified recalling the unichromatic reflection model, from equation (17), considering narrow sensor bands:

$$R = m_b(\vec{n},\vec{s})e(\lambda_R)c_b(\lambda_c) \tag{28}$$

With a change in illuminant color from $e(\lambda)$ to $e_1(\lambda)$, we have

$$R_1 = m_b(\vec{n},\vec{s})e_1(\lambda_R)c_b(\lambda_c) \tag{29}$$

From equations (28) and (29),

$$R_1 = \left[e_1(\lambda_R) / e(\lambda_R) \right] R \tag{30}$$

From equation (30) it can be seen that a change in the color of the illuminant affects the response of each color sensor by a corresponding scalar factor. Therefore, the new mean values of the red, green, and blue bands in the image become $\alpha\overline{r}$, $\beta\overline{g}$, and $\gamma\overline{b}$, where α, β, and γ are scalars. Considering that, under the new illumination $e_1(\lambda)$, the scalars α, β, and γ are present in both numerator and denominator of equations (25)-(27), the dependence on illumination color is removed.

The color constancy procedure iteratively performs these two types of normalization until the dissimilarity between two successive resultant images is below an acceptance level. It is possible to demonstrate that the technique converges and provides unique results [8].

The results show an improvement in the stability of saliency, because of the invariance to illuminant color (see Section 5), but at the expense of a significantly higher computational cost, due to the iterative nature of the involved computation. This technique has also the drawback of a high sensitivity to changes in viewpoint and to the inclusion of new objects in the scenes, because of its dependence on the global color composition of the images.

4 A new approach: visual saliency using color ratios

This section describes the proposal of a new visual saliency algorithm with embedded color constancy properties.

With the purpose of obtaining contour images with good color constancy properties, Gevers and Smeulders [10] developed the color space $m_1m_2m_3$, based on the color ratio between neighboring image pixels (x_1, x_2):

$$m_1 = R^{x_1}G^{x_2} / G^{x_1}R^{x_2} \tag{31}$$

$$m_2 = R^{x_1}B^{x_2} / B^{x_1}R^{x_2} \tag{32}$$

$$m_3 = G^{x_1}B^{x_2} / B^{x_1}G^{x_2} \tag{33}$$

This differential version of color constancy gave us the idea of generalizing the concept of gradient between neighboring pixels to that of center-surround opposition. Thus, invariance of color gradients would turn into our desired invariance of center-surround oppositions. Under this approach, the x_1 pixel is replaced by the center region and the x_2 pixel by the surround region. Moreover, the ratios do no longer relate color bands, but color opponents (see equations 1-4), as follows:

$$RG = R_o^c G_o^s / G_o^c R_o^s \tag{34}$$

$$GR = R_o^s G_o^c / G_o^s R_o^c \tag{35}$$

$$BY = B_o^c Y_o^s / Y_o^c B_o^s \tag{36}$$

$$YB = B_o^s Y_o^c / Y_o^s B_o^c \tag{37}$$

where $R_o^c, G_o^c, B_o^c \text{ and } Y_o^c$ are opponent red, green, blue and yellow at center regions and $R_o^s, G_o^s, B_o^s \text{ and } Y_o^s$ are opponent red, green, blue and yellow at surround regions. The *RG* opponency corresponds to a visual field that is excited by red stimuli in the center and by green stimuli in the surround, and inhibited by red stimuli in the surround or green stimuli in the center. The *GR* corresponds to the converse. The same consideration is valid for the blue and yellow color pair. With the use of centers and surrounds at different scales, located at coarser or finer levels in the Gaussian pyramids, it is possible

to compute the color opponencies at multiple scales, according to the visual saliency model described in Section 2.3.

Assuming that neighboring center and surround regions have a locally constant illuminant, the same surface normal and uniform albedo, according to the unichromatic reflection model, from equations (17) and (34), we have:

$$RG = \frac{(m_b^c(\vec{n},\vec{s})\ e^c(\lambda_R)\ c_b^c(\lambda_R))(m_b^s(\vec{n},\vec{s})\ e^s(\lambda_G)\ c_b^s(\lambda_G))}{(m_b^s(\vec{n},\vec{s})\ e^s(\lambda_R)\ c_b^s(\lambda_R))(m_b^c(\vec{n},\vec{s})\ e^c(\lambda_G)\ c_b^c(\lambda_G))} = \frac{c_b^c(\lambda_R)c_b^s(\lambda_G)}{c_b^s(\lambda_R)c_b^c(\lambda_G)} \qquad (38)$$

which is only dependent on the sensors and the surface albedo. The same can be done for equations (35)-(37). A key feature of the color ratios presented in equations (34) to (37) is their invariance to both intensity and color normalizations, which makes them intrinsically invariant to lighting intensity and illumination color changes. Moreover, the ratios have a local nature, thus avoiding the distorting effects possibly introduced by global normalizations.

It is important to observe that, in the two preprocessing approaches (Section 3), the saliencies were proportional to the value differences between center and surround regions, while here they are proportional to the value ratios of these regions. We use the logarithms of the spaces (R_o/G_o) and (Y_o/B_o), so that we can compute the opponencies by differences of logarithms across the scales, instead of divisions. Additionally, as the logarithm of the inverse of an expression is the negative logarithm of the expression, we have only two pyramids for color, one for $\ln(R_o/G_o)$ and another for $\ln(Y_o/B_o)$.

The logarithms of the quotients are computed as differences of logarithms, and the individual $\ln(R_o)$, $\ln(G_o)$, $\ln(B_o)$ and $\ln(Y_o)$ values are saturated having the unity as minimum value, in order to avoid instability and negative values. Moreover, the masking of low-intensity pixels commented in the previous sections applies also here.

The partial saliency maps from each center-surround scale and opponent color combination are normalized through exponentiation before combining them, restoring the linear proportion between the partial maps.

Summarizing, the proposed multiscale color ratio algorithm consists of the following steps:

1. Conversion from input *RGB* space to opponent color space $R_oG_oB_oY_o$, using equations (1) to (4).
2. Construction of the $\ln(R_o/G_o)$ and $\ln(Y_o/B_o)$ Gaussian pyramids, with 8 scale levels.
3. Computation of the multiscale color ratios through differences of logarithms at pyramid center levels $c \in \{2,3,4\}$ and their corresponding surround pixels at levels $s = c + d$, $d \in \{3,4\}$, according to equations (34)-(37) and Section 2.3.
4. Generation of the resultant saliency map as the sum of the partial maps subject to exponentiation and weighting according to their entropy content (Section 2.3.4).

5 Performance comparison

In order to assess the relative performance of the algorithms, we made qualitative and quantitative analysis of the saliency results for images of the same scenes subject to different illumination conditions, and also compared execution times.

5.1 Qualitative analysis of results

We have compiled the experimental results for three scenarios in Figures 2-4. The results corresponding to lighting intensity normalization are shown in the second

columns of such figures, comprehensive color normalization in the third columns and multiscale color ratios in the fourth columns.

In Figure 3, under *lighting intensity normalization* (second column), the most salient regions change from the gravel path in the first image, to the green areas at the left and at the center in the second image, and to the red roof at the left, orange flowers at the right, and central green areas in the third image. In the three images, part of the reddish bushes at the left were marked as salient. In sum, the detection of salient regions is very sensitive to the illumination changes.

Under *comprehensive color normalization* (third column), the salient regions in all images correspond to the tree in the horizon line at the left, the orange flowers at the right, and the reddish bushes, although the saliency peaks change from the tree at horizon line in the first image, to the orange flowers in the second image, and to the orange flowers and the green area at the left in the third image. Part of the reddish bushes at the left were again marked as salient in the three images. We observe thus that salient regions are more stable than in the former case, although the most salient one changes from image to image.

With *multiscale color ratios* (fourth column), in all images the red roof, the orange flowers and the reddish bushes are identified as the most salient regions, although in the last image the houses in the center do also appear as salient.

In Figure 4, in addition to illumination changes, the four images were taken at slightly different points of view. Under *lighting intensity normalization* (second column), for the first image, the most salient regions correspond to the yellowish bushes at the left and right sides, and to the top of the trees in the center. For the subsequent

images, the most salient region changes to the yellowish bush at the right, then to the yellowish bush at the left.

Under *comprehensive color normalization* (third column), the same salient regions as in the former normalization were identified in the four images. In addition, the red house has a more accentuated saliency in the second and fourth images.

With *multiscale color ratios* (fourth column), in all images the red house is stably identified as a salient region. In the last three images, the yellowish bushes are pointed as salient in the same manner as in the two former color normalizations.

In Figure 5, under all color normalizations and in all images, the yellow flowers are indicated as salient, while the reddish tree at the left is not always indicated as salient, the *comprehensive color normalization* giving the most stable results for it. Note that the third image is saturated, and the non-linear distortion affects more heavily the normalization based on color ratios, resulting in the reddish tree not being indicated as salient as in the previous two images.

In general, it can be observed that the stability of the saliency maps obtained using *lighting intensity normalization* is poor. The saliencies obtained with *comprehensive color normalization* are outstandingly more stable across the images. However, since comprehensive color normalization uses averages of color components over the entire image, the inclusion of new salient regions affects the overall saliency more than in the case of intensity normalization. This effect can be observed in the third image in Figure 4, where the red house is no longer significantly salient.

Using the *multiscale color ratio* approach, a better stability than with *lighting intensity normalization* is observed, while the results are qualitatively similar to those obtained with *comprehensive color normalization*.

The ratio nature of the *multiscale color ratios* approach results in saliency images where the salient areas are much stronger than the background, facilitating the subsequent task of segmentation used to isolate the salient regions for further characterization.

5.2 Quantitative analysis of results

In the color literature, a common image comparison measure is the root mean squared error (RMSE) [1, 7]. In our experimentation, the first image of each scene was selected as reference, and its resulting saliency map was compared using RMSE with the saliency maps of the other images of the same scene subject to different illuminations. With the aim of assessing the sensitivity of the different algorithms to the illumination changes, especially into what concerns the extraction of salient cues, we made a sequence of RMSE measures taking into account only the most salient pixels in the saliency maps, within a range from 1% to 100% of them.

The behaviors of the three algorithms are displayed in Figure 6. The *multiscale color ratios* approach presents the lowest RMSE values for all the scenes and illumination changes, and it is also the approach less sensitive to what fraction of the pixels is selected. The first indicates a better stability against illumination changes and the second indicates that the changes of the less-salient pixels are not significant to the overall saliency output. This effect is partially due to the concentration of saliency output in the most salient pixels of the source image.

The two approaches based on color constancy preprocessing have significantly greater RMSE when a small percentage of the most salient pixels are selected. Since our objective is to identify the most salient regions in the images and take them as landmark candidates, the superiority of the *multiscale color ratios* algorithm for this task seems

clear. The maximum advantage is observed when about 10% to 30% of the most salient pixels are selected.

It should be observed that the most significant RMSE results are those obtained with few of the most salient pixels, e.g. less than 20% of them, because the reduced background intensity of the saliency maps using the *multiscale color ratios* algorithm is favorable to it when comparing background regions.

5.3 Execution time analysis

To compare and analyze the execution times we made a benchmark evaluation, applying the three color constancy techniques to a source *RGB* image of 512x384 pixels averaging 100 successive executions of each approach. Table 1 shows the execution times obtained, using a standard PC computer (AMD Athlon 800MHz, 128Mb DRAM, Windows 98). It can be observed that saliency detection with our multiscale color ratio technique needs lower execution time than the other approaches. The reason for this is discussed in what follows.

To construct each Gaussian pyramid the separability of Gaussian filtering is used, which permits its efficient implementation using successive horizontal and vertical convolutions with a 5-tap filter mask (with weights 1, 4, 6, 4, and 1). The number of pixels in a Gaussian pyramid converges to N*4/3 pixels with the increment of the number of levels, where N is the number of pixels in the original image. Then, to fill one Gaussian pyramid it is necessary to execute the convolution for N*4/3 pixels, and each pixel requires two passes of the 5-tap mask, resulting in N*32/3 float additions and N*40/3 float multiplications.

Table 2 indicates the number of operations required to process an image for the three evaluated algorithms. The proposed multiscale color ratio requires fewer operations

than the other approaches, mainly due to the unification of the color constancy and saliency detection processes. Data manipulation operations are not considered in this comparison, since they are dependent on the implementation. Table 3 shows the distribution of computing time between the most important tasks carried out by the proposed algorithm. The computation of center-surround differences is only 8% of the total execution time, because these differences are computed at the scale of the centers, instead of at the source image scale. For example, for a 512x384-pixel image, the center-surround differences between levels 3 and 7 are computed using the dimensions of the center image at level 3 of the pyramids, i.e., 64x48 pixels.

6 Conclusions

In this paper, we have compared three approaches to color constancy as applied to a landmark detection system based on opponent-color saliency.

The first approach, lighting intensity normalization through the transformation of color from *RGB* to chromaticity space, has shown an undesirable sensitivity to shadows and changes in the illuminant color and viewpoint.

The comprehensive color normalization has proven to be more stable to illumination changes than the lighting intensity normalization, but presents higher computational cost and also produces undesired changes in the detected salient regions. The color constancy is affected by the global color measures in the image, and so the technique is sensitive to the inclusion/exclusion of objects in the scenes.

The proposed color ratios constitute direct measures of color opponencies, which are intrinsically invariant to both lighting intensity and illuminant color changes. Additionally, their definition based on local features makes them resistant to moderate

viewpoint changes. Moreover, all this is attained at a lower computational cost than with the two previous approaches.

We conclude that, for the target application, i.e., detecting salient visual cues for tentative landmark extraction, our technique is more suitable than the other two, because it provides more stable results in scenarios subject to illumination changes, like those occurring in outdoor environments.

Acknowledgments

The authors would like to thank Enric Celaya for comments about this paper, and the support obtained from the *Forschungszentrum Informatik* and *Institut für Prozessrechentechnik, Automation und Robotik*, Karlsruhe University, Germany. This work is partially supported by the Spanish Science and Technology Directorate, in the scope of the project "Reconfigurable system for vision-based navigation of legged and wheeled robots in natural environments (SIRVENT)", grant DPI2003-05193-C02-01.

7 References

[1] K. Barnard, Modeling scene illumination color for computer vision and image reproduction: a survey of computational approaches, Ph.D. Thesis, Computer Science Dept., Simon Fraser University, Burnaby, Canada, 1998.

[2] J. Berens and G. D. Finlayson, Log-opponent chromaticity coding of colour space, in: Proceedings of International Conference on Pattern Recognition (ICPR 2000), Vol. 1, 2000, pp. 206-211.

[3] V. Bruce, P. R. Green, and M. A. Georgeson, Visual Perception, Psychology Press, United Kingdom, 1997.

[4] P. J. Burt, The pyramid as a structure for efficient computation, in: A. Rosenfeld, Ed., Multiresolution image process and analysis, Springer-Verlag, Berlin-Heidelberg, 1984, pp. 6-35.

[5] G. D. Finlayson, M. S. Drew, and B. Funt, Spectral sharpening: sensor transformations for improved color constancy, Journal of Optical Society of America 11 (1994) 1553-1563.

[6] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, Color by correlation, Proceedings of 5th Colour Imaging Conference, 1997, pp. 6-11.

[7] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, Color by correlation: a simple, unifying framework for color constancy, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 1209-1221.

[8] G. D. Finlayson, B. Schiele, and J. L. Crowley, Comprehensive colour image normalization, in: Proceedings of 5th European Conference on Computer Vision (ECCV '98), Freiburg, Germany, 1998, pp. 475-490.

[9] M. M. Fleck, D. A. Forsyth, and C. Bregler, Finding naked people, in: Proceedings of 4th European Conference on Computer Vision, Cambridge, 1996, pp. 593-602.

[10] T. Gevers and A. W. M. Smeulders, Color-based object recognition, Pattern Recognition 32 (1999) 453-464.

[11] J. P. Gottlieb, M. Kusunoki, and M. E. Goldberg, The representation of visual salience in monkey parietal cortex, Nature 39 (1998) 481-484.

[12] L. Itti, C. Koch, and E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (1998) 1254-1259.

[13] R. Jain, R. Kasturi, and B. G. Schunck, Machine Vision, McGraw-Hill, New York, 1995.

[14] M. E. Jernigan and F. D'Astous, Entropy-based texture analysis an the spatial frequency domain, IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (1984) 237-244.

[15] T. Kadir and M. Brady, Saliency, scale and image description, International Journal of Computer Vision 45 (2001) 83-105.

[16] G. J. Klinker, S. A. Shafer, and T. Kanade, A physical approach to color image understanding, International Journal of Computer Vision (1990) 7-38.

[17] R. Murrieta-Cid, M. Briot, and N. Vandapel, Landmark identification and tracking in natural environment, LAAS, Toulouse, France, Report 98037, 1998.

[18] K. Nagao and W. E. L. Grimson, Using photometric invariants for 3D object recognition, Computer Vision and Image Understanding 71 (1998) 74-93.

[19] H.-C. Nothdurft, Saliency from feature contrast: additivity across dimensions, Vision Research 40 (2000) 1183-1201.

[20] Y. Otha, T. Kanade, and T. Sakai, Color information for region segmentation, Computer Graphics and Image Processing 13 (1980) 222-241.

[21] S. J. Sangwine and R. E. N. Horne, The color image processing handbook, 1st ed, Chapman & Hall, London, 1998.

[22] S. Santini and R. Jain, Gabor space and the development of preattentive similarity, in: Proceedings of International Conference on Pattern Recognition, Vienna, Austria, 1996.

[23] C. E. Shannon, A mathematical theory of communication, The Bell System Technical Journal 27 (1948) 379-423.

[24] M. J. Swain and D. H. Ballard, Color indexing, International Journal of Computer Vision 7 (1991) 11-32.

[25] T. S. C. Tan and J. Kittler, Colour texture analysis using colour histogram, IEEE Vision Image and Signal Processing 141 (1994) 403-412.

[26] E. Todt and C. Torras, Detection of natural landmarks through multiscale opponent features, in: Proceedings of 15th International Conference on Pattern Recognition, Barcelona, Spain, IAPR, Vol. 3, 2000, pp. 976 - 979.

[27] A. M. Treisman and G. Gelade, A feature-integration theory of attention, Cognitive Psychology 12 (1980) 97-136.

[28] R. VanRullen, Visual saliency and spike timing in the ventral visual pathway, Journal of Physiology - Paris (2002).

[29] G. Wyszecki and W. S. Stiles, Color Science: Concepts and Methods, Quantitative Data and Formulae, John Wiley and Sons, U.S.A., 1982.

[30] K. Yamaba and Y. Miyake, Color character recognition method based on human perception, Optical Engineering 32 (1993) 33-40.

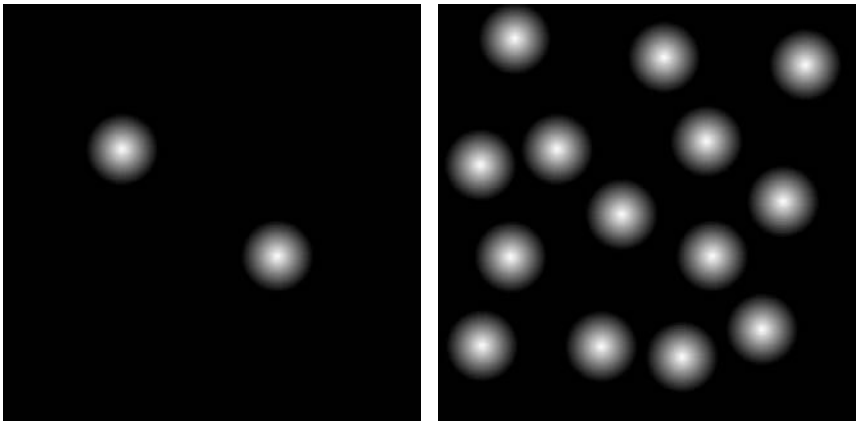
About the authors

Eduardo Todt received a BS degree in Electrical Engineering and a MS degree in Computer Science from the Universidade Federal do Rio Grande do Sul, Brazil, in 1985 and 1990, respectively. In 1989 he became an assistant professor in the Computer Science Faculty of the Pontifícia Universidade Católica do Rio Grande do Sul, Brazil, and currently he is carrying out his PhD at Universitat Politècnica de Catalunya, Spain. His major research interests are Computer Vision and Industrial Automation.

Carme Torras (http://www-iri.upc.es/people/torras) is Research Professor at the Spanish Scientific Research Council (CSIC). She received M.Sc. degrees in Mathematics and Computer Science from the Universitat de Barcelona and the University of Massachusetts, respectively, and a Ph.D. degree in Computer Science from the Universitat Politècnica de Catalunya. Prof. Torras has published three books and more than a hundred papers in the areas of Robotics, Vision, and Neurocomputing. She has been local project leader of several European projects, such as ``Planning RObot

Motion" (PROMotion), ``Behavioural Learning: Sensing and Acting" (B-LEARN), ``Robot Control based on Neural Network Systems" (CONNY) and ``Self-organization and Analogical Modelling using Subsymbolic Computing" (SUBSYM).

List of figures



H=0.65 *H*=3.53

Figure 1 - Relevance of saliency maps is affected by overall saliency present in each map. In the left map, there are only two saliency peaks, while in the right map there are 14 identical saliency peaks. Although the saliency peaks have the same value, the conspicuity of the peaks in the left map is larger than that in the right map. The lower entropy of the left image is consistent with the lower dispersion of the saliency spots.

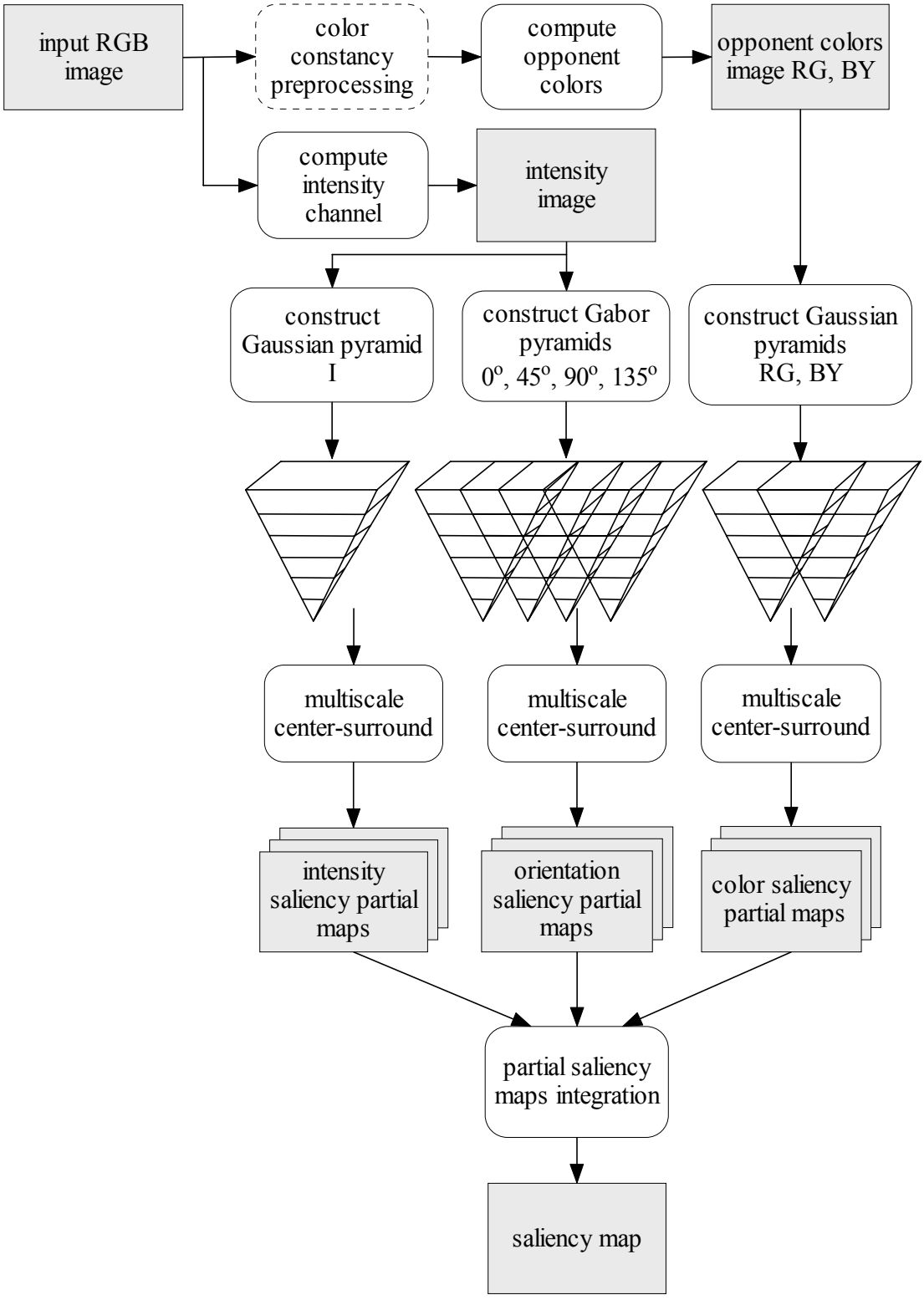


Figure 2 - Diagram of the visual saliency detection system. The dashed module is included only in the two approaches relying on color constancy preprocessing described in Section 3.

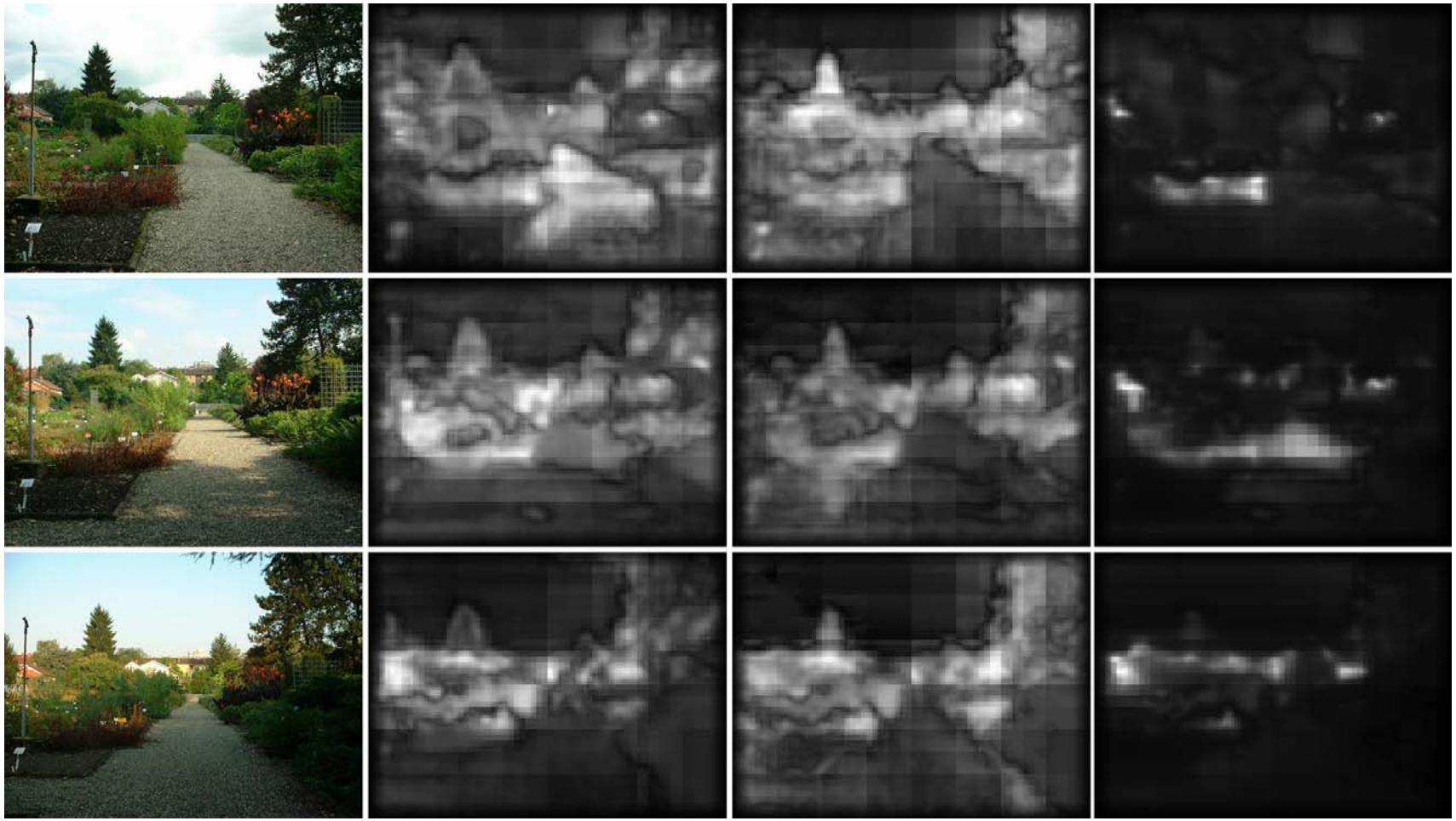


Figure 3 - Saliency maps computed for scene “A” for three different illumination conditions (one in each row). Each source image (left column) was processed with lighting intensity normalization (second column), lighting intensity and illuminant color normalization (third column), and color ratios (fourth column). The whiter regions indicate the more salient parts detected.

Figure 4 - Saliency maps computed for scene “B” for four different illumination conditions (one in each row). Note that there are slight changes in perspective between the images. Each source image (left column) was processed with lighting intensity normalization (second column), lighting intensity and illuminant color normalization (third column), and color ratios (fourth column). The whiter regions indicate the more salient parts detected.

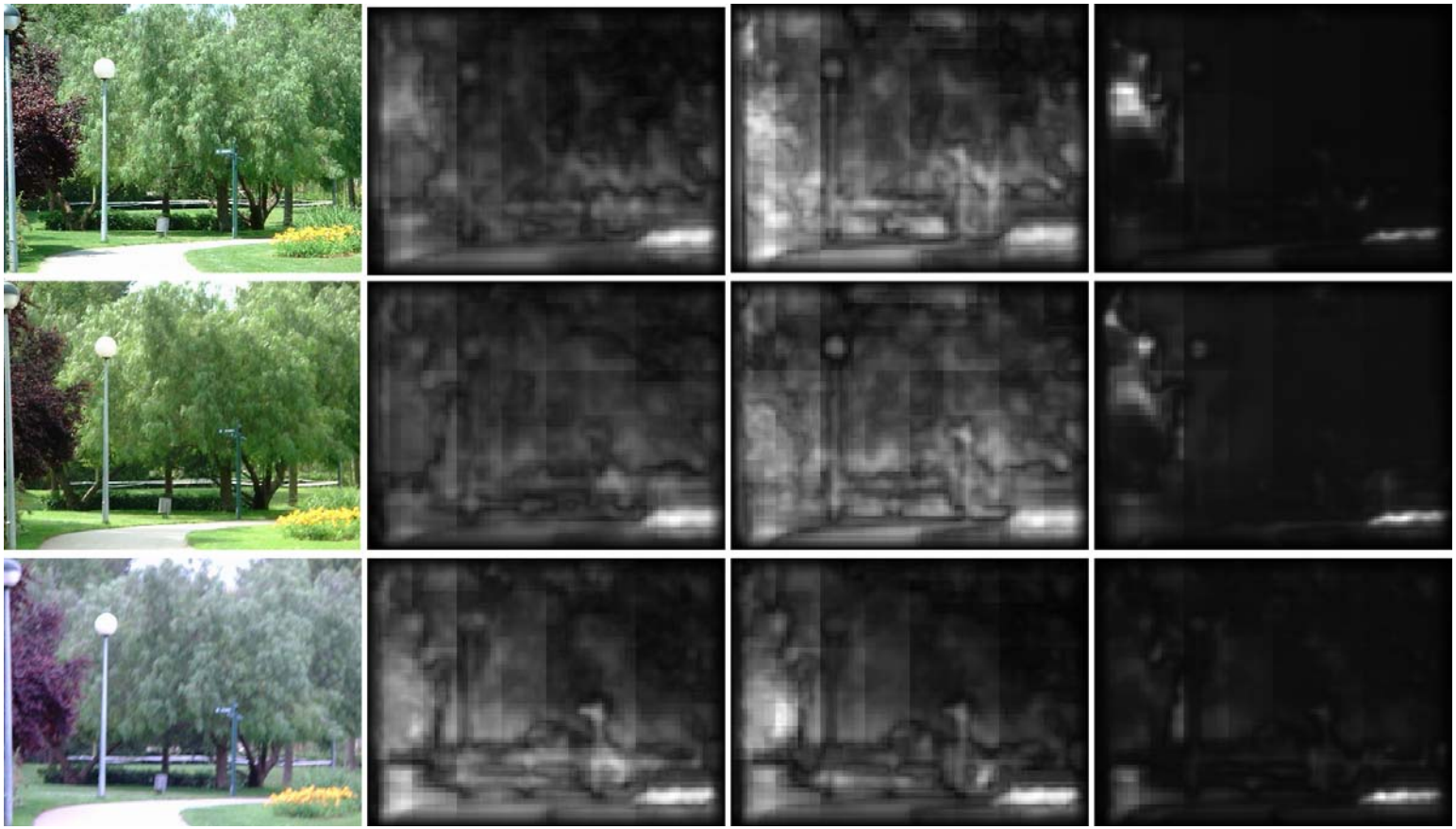


Figure 5 - Saliency maps computed for scene “C” for three different illumination conditions (one in each row). Each source image (left column) was processed with lighting intensity normalization (second column), lighting intensity and illuminant color normalization (third column), and color ratios (fourth column). The whiter regions indicate the more salient parts detected.

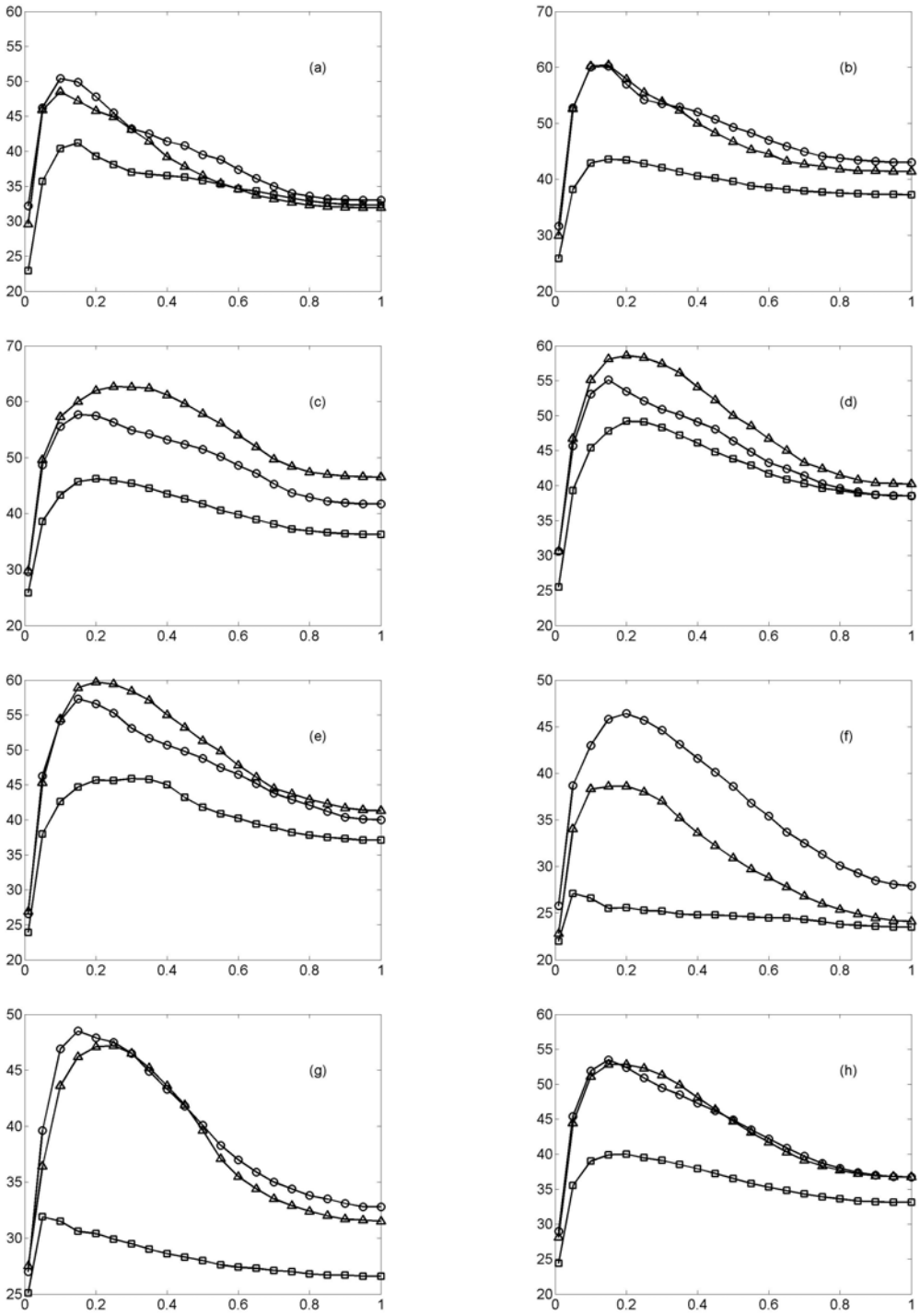


Figure 6– RMSE between the saliency maps corresponding to pairs of images of the same scene under different illuminations. The abscissas axis indicates the percentage of most salient pixels considered. The curves with circles, triangles, and squares refer to intensity normalization, comprehensive color normalization, and multiscale color ratios, respectively. Graphs (a) and (b) correspond to Figure 3; (c), (d), and (e) to Figure 4; (f) and (g) to Figure 5; and (h) represents the mean RMSE of all images. The first image in each set is taken as the reference image, against which the other images are compared.

List of tables

Table 1 - Execution times for computing visual saliency with the three approaches studied.

Approach	Seconds
Intensity normalization	0.86
Comprehensive color normalization	1.19
Multiscale color ratio	0.77

Table 2 - Amount of floating-point operations per pixel performed by the three evaluated approaches to visual saliency based on different color constancy techniques. Values are calculated for 8-level pyramids, and center-surround differences at levels 2-5, 3-6, and 4-7. Moreover, three iterations of comprehensive color normalization are assumed.

	Float additions	Float subtractions	Float multiplications	Float divisions	Float logarithms
Intensity normalization					
RGB to Intensity and rgb	2.0			3.0	
rgb to RGBY	3.0	6.0		3.0	
4 x Gaussian pyramids	42.7		53.3		
center-surround differences	0.5	0.5			
TOTAL	48.2	6.5	53.3	6.0	0.0
Comprehensive color normalization					
3x intensity normalization	6.0			9.0	
3x lighting color normalization	9.0			9.0	
rgb to RGBY	3.0	6.0		3.0	
4 x Gaussian pyramids	42.7		53.3		
center-surround differences	0.5	0.5			
TOTAL	61.2	6.5	53.3	21.0	0.0
Multiscale color ratio					
RGB to RGBY	3.0	6.0		3.0	
ln of RGBY					4.0
ln(R)-ln(G), ln(Y)-ln(B)		2.0			
2 x Gaussian pyramids	21.3		26.7		
center-surround differences	0.5	0.2			
TOTAL	24.8	8.2	26.7	3.0	4.0

Table 3 - Distribution of execution time between the tasks performed within the multiscale color ratio approach.

Task	Fraction of total execution time
Conversion *RGB* to *R'G'B'Y'*	0.21
Logarithm of *R'G'B'Y'*	0.22
Pyramids ln(*R'*/*G'*), ln(*Y'*/*B'*)	0.23
Center-surround differences	0.08
Other tasks	0.26