Preattentive Colour Features by Steerable Filters*

Udo Mahlmeister**, Bertil Schmidt, Gerald Sommer

Institut für Informatik
Christian–Albrechts–Universität zu Kiel
Preußerstrasse 1–9, D–24105 Kiel, Germany
Tel.: (0431) 56 04–33, Fax: (0431) 56 04–81
email: uhm@informatik.uni-kiel.de

Abstract. Visual search is the task of finding objects in an image which are described in a high dimensional space, spanned by preattentive features, e.g. orientation, scale, and colour. By using steerable filters this search space may be scanned continouosly, though spanned by discrete feature detectors. Based on this idea, we will show a method for detecting arbitrarily oriented bars from steerable filter responses. Detection performs rather invariant to illumination colour, exploiting the properties of the CIE–*Lab* colour space.

1 Introduction

1.1 Selective Attention in Artificial Visual Systems

Visual attention is the capability of a visual system to sequentially focus its processing resources to a selected part of the visual field. This selection process is guided by the degree of interest of image locations with respect to a current visual task. The degree of interest is defined in terms of few local feature dimensions, say scale, orientation, motion, texture and colour. These *preattentive features* are supposed to reflect the most important aspects of object surfaces and boundaries. According to the basic model of visual attention by Koch and Ullman[10], the *focus of attention (FOA)* is moved to the maximum of the *saliency map* which is a task guided combination of preattentive feature maps. The degree of interest assigned to an image structure may then be controlled by the weighting of the preattentive feature maps. This processing scheme is supposed to reduce the complexity of a visual task[14].

The spatial coordinates of an image structure may be considered as two additional feature dimensions in the preattentive feature space. Then the task of finding all structures in an image with a specific feature combination in specific spatial relations means to perform a full search in this space. Clearly, the complexity of this search explodes as the number of feature dimensions increases. Despite this, in an attention architecture the search is broken down into two stages. In the first preattentive stage a search is performed for a fixed feature combination in the whole visual field. The dimensionality of the search space is reduced to two spatial coordinates. The result of this spatially parallel process is a set of potentially interesting locations. In the second, attentive stage the suggested locations are attended and are analyzed within the orthogonal complement of the spatial subspace. This way the effort of searching along the spatial dimensions is shifted to time. The information flow in this processing scheme is not strictly linear. The features registered at one FOA may influence the guess for the next FOA. Hereby its goal is to maximize the evidence for a

* in Mustererkennung 1995, 17. DAGM–Symposium, Bielefeld, Springer–Verlag
** partially supported by DFG, grant So 320/1-1

known object[7] or, equivalently, to minimize the number of FOA necessary in order to recognize an object or situation. Also, the behaviour of the system may be influenced by long time experience. Applied to the task of face recognition, the eyes, the nose, and the mouth of a face are to be located. As recently was shown by Herpers[8], the second eye in a face image is located more easily, if its position and features are predicted by features of the first eye.

The neurological grounding of natural attention systems becomes apparent at the the inhomogeneous receptor distribution of the mammalian retina. While the periphery of the retina covers the whole visual field with rather low spatial and temporal resolution, its central part, the fovea, offers high resolution although in a rather small area. This way the preattentive task of localizing interesting points is supported mainly by the periphery. On the other hand, the feature analysis in order to recognize objects needs the high resolution of the fovea.

The second main advantage of attentional strategies happens to be important in general purpose vision architectures. Since behaviour is the interaction with objects, most vision tasks are set in an object related form, and objects are mostly described by means of surface properties. Despite this, at an early representation level objects are represented as spatio–temporal feature distributions directly derived from the 2D–sensory input. The attention mechanism serves as the glue between these two description levels. It allows object surfaces and boundaries to be described by combinations of preattentive features. The geometry of objects is inferred from the scan path, i. e. the sequence of FOA coordinates. Furthermore, if tracking is considered as a special form of attention, object motion may be derived directly from the camera movement signal of a tracking system.

The linking between high and low–level information can only be achieved, if preattentive features correspond with object surface properties. Therefore surface properties have to be extracted *invariant* with respect to some environmental conditions, e.g. perspective transformations or lighting.

1.2 Colour Constancy

Probably the most important surface property is colour. Because humans are able to roughly discriminate between spatial variations in illumination and variations in surface colour, they are able to perceive the colour of a surface rather independent from the prevailing illumination. This competence, termed as *colour constancy* enables humans to use colour as a feature to distinguish many objects in a large variety of environments.

The early stages of the Human Visual System employ a multiscale multi orientation representation of the sensory input. It is commonly agreed, that this kind of representations are also appropriate for the majority of computer vision problems [3]. We will exploit the properties of these representations for the colour invariant extraction of preattentive features. Though orientation selectivity is an essential property of these scheme, it is not used in other approaches to colour constancy. Neither Horns classical retinex–based algorithm[9, 11], nor more recent colour constant colour indexing techniques[6] use the information of orientation.

1.3 Visual Search in Continuous Feature Spaces by Steerable Filters

Visual search is the most important subtask of visual attention, since it is an inherent part of many vision problems. It is also the most employed task in psychological experiments to examine visual attention[13]. A typical test pattern in these experiments contains coloured rectangles on a homogenouos background (Fig. 2a) which are completely specified by their colour, size, and orientation. A pattern also shows a single search target, which differs from the surrounding distractors in at least one feature. The time it takes for the subject to detect the target is measured for a varying number of distractors. If the target is uniquely distinguished from all distractors by at least one feature, the search time does not depend on the number of distractors. Therefore, the task is assumed to be done in parallel. Otherwise, if the target shares some features with distractors, i.e. it is uniquely described only by a conjunction of features, the search time usually grows linearly with the number of distractors. In this case the task is thought of requiring a sequential scan of all objects, until the target is found. This procedure, described in Treisman's feature integration theory[13] may go back to architectural constraints of the human visual system, but bears no obvious computational advantages for an artificial system. Therefore we consider visual search in the concrete form as "fast conjunctive search" i.e. the task of finding all objects specified by a conjunction of at least one preattentive feature in parallel.

The attention system we use to demonstrate our procedure is based on Ahmad's VISIT[1] which was designed to simulate visual search experiments described above. Objects are localized by a maximum detection on the saliency map which is a weighted sum of the output of linear filters at several scales and orientations on three colour signals. Of coarse, a target is most easily detected, if the weights match the objects description in the feature space of scale, orientation, and colour. Referring to orientation, the weight of the filter with the desired orientation is set to one, while all other weights are set to zero. Two questions arise in this context. First, how are the weights adjusted, if the targets orientation doesn't match with one of the filters orientation? Secondly, how are the weights adjusted, if the orientation of the target is not specified by the search task? The first question is answered most naturally by the theory of *steerable filters*[4] which provides a calculus to synthesize an arbitrarily oriented filter by the linear combination of some fixed set of basis filters. This way any orientation could be searched for by the small effort of interpolating few filter responses instead of applying the matching filter to the whole image. The second problem will be solved readily, if there exists a linear combination of this basis filter set which synthesizes an isotropic filter. On this condition equal filter outputs for all orientations are guaranteed. We will use Andersson filters[2] which ideally meet both conditions.

In the next section we will provide a computational model for colour invariant preattentive feature extraction and integration based on the responses from a set of steerable filters. The assumptions underlying colour constancy are described in section 3. Finally, we will give some preliminary experimental results in section 4.

2 The Feature Integration Process

Given an input image, the following three step algorithm (Fig. 1) yields a saliency map according to the prespecified target feature combination.

Step 1: approximates the nonlinear characteristic of colour sensors by converting the sensory input from RGB– to Lab–colour space.
Step 2: provides the basis filter outputs by convolving Lab–images with a set of scaled and rotated Andersson filters.
Step 3: combines these these linearly outputs to synthesize a filter matching to the specified feature conjunction.

2.1 Colour Space Conversion

The triple $(r_R(\boldsymbol{x}), r_G(\boldsymbol{x}), r_B(\boldsymbol{x}))$ denotes the tristimulus values of the input colour image as a function of the spatial coordinate vector $\boldsymbol{x} = (x, y)^T$. This tristimulus is converted to the well known CIE–Lab–colour space by:

$$q_L = 25\ (100\, q_Y/q_{Y0})^{\frac{1}{3}} - 16 \qquad \text{if} \quad \frac{1}{10} \leq q_Y/q_{Y0} \leq 1 \tag{1}$$

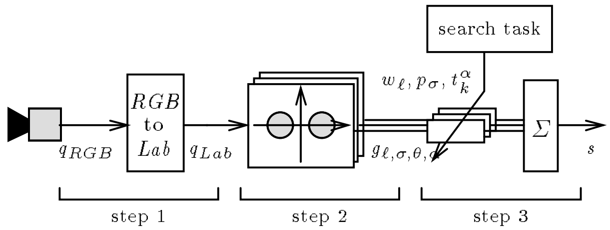


Fig. 1. Three step algorithm: colour space conversion, linear filtering and feature integration.

$$q_a = 500 \left[(100\, q_X/q_{X0})^{\frac{1}{3}} - (100\, q_Y/q_{Y0})^{\frac{1}{3}} \right]$$
$$q_b = 200 \left[(100\, q_Y/q_{Y0})^{\frac{1}{3}} - (100\, q_Z/q_{Z0})^{\frac{1}{3}} \right]$$

where $(q_X(\boldsymbol{x}), q_Y(\boldsymbol{x}), q_Z(\boldsymbol{x}))$ are the nonnegative CIE–XYZ tristimulus values, obtained by a linear mapping from the NTSC–RGB values:

$$\begin{pmatrix} q_X \\ q_Y \\ q_Z \end{pmatrix} = \begin{pmatrix} 0.607 & 0.174 & 0.201 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.117 \end{pmatrix} \begin{pmatrix} q_R \\ q_G \\ q_B \end{pmatrix} \tag{2}$$

The values (q_{X0}, q_{Y0}, q_{Z0}) denote the reference white.

2.2 Filtering and Steering Orientation

Each component of the Lab–colour–signal $q_\ell,\ \ell \in \{L, a, b\}$ is now convolved $(*)$ with a set of Andersson basis filters tuned to specific scales $\sigma_i = \sigma_b 2^i,\ i = 0, 1, 2,$ orientations $\theta_k = k\pi/4,\ k = 0, \ldots, 3$ and phases $\alpha \in \{even, odd\}$:

$$\begin{pmatrix} g_{L,\sigma,\theta,\alpha} \\ g_{a,\sigma,\theta,\alpha} \\ g_{b,\sigma,\theta,\alpha} \end{pmatrix} = \begin{pmatrix} q_L \\ q_a \\ q_b \end{pmatrix} * h_{\sigma,\theta,\alpha} \tag{3}$$

The odd and even symmetric versions $g_{\ell,\sigma,\theta,e}(\boldsymbol{x})$ and $g_{\ell,\sigma,\theta,o}(\boldsymbol{x})$ constitute a Hilbert pair. Since the Andersson filter is defined in the frequency domain, we will consider its transfer function $H_{\sigma,\theta,\alpha}(u, v)$. Rewritten in polar frequencies $(\rho, \vartheta) = (\sqrt{u^2 + v^2}, \arg(u, v))$ the Andersson transfer function becomes separable: $H_{\sigma,\theta,\alpha}(\rho, \vartheta) = R_\sigma(\rho) \cdot A_{\theta,\alpha}(\vartheta)$ Its radial component is:

$$R_\sigma(\rho) = \cos^2 \frac{\pi \ln(\sigma\rho)}{2B \ln 2} \quad \text{für} \quad 2^{-B}\sigma^{-1} \le \rho < 2^B \sigma^{-1} \tag{4}$$

with a constant logarithmic bandwidth B. The angular component of the first order Andersson filter is defined as:

$$A_{\theta,\alpha}(\vartheta) = \begin{cases} \cos^2(\vartheta - \theta) & \text{if } \alpha = even \\ j\, \cos^2(\vartheta - \theta)\, \mathrm{sgn}(\cos(\vartheta - \theta)) & \text{else} \end{cases} \tag{5}$$

The Andersson filter is steerable[2] with respect to orientation θ, since the angular component to an arbitrary orientation can be written as a linear superposition of rotated copies of itself:

$$A_{\theta,\alpha}(\vartheta) = \sum_{k=0}^{M-1} b_k^\alpha(\theta)\, A_{\theta_k,\alpha}(\vartheta) \tag{6}$$

The minimum number $M_0 \le M$ of basis functions to steer the function $A_{\theta,\alpha}(\vartheta)$ is equal to the number of nonzero coefficients a_n of its Fourier expansion:

$$A_{\theta,\alpha}(\vartheta) = \sum_{n=-N}^{N} a_n e^{jn\vartheta} \tag{7}$$

While the even symmetric component $A_{\theta,e}(\vartheta)$ needs 3 basis functions, the odd component $A_{\theta,o}(\vartheta)$ corresponds to an infinite series and would require infinitely many basis functions. A finite series for $n = -3, \ldots, 3$ with four nonzero coefficients yields a good approximation of $A_{\theta,o}(\vartheta)$. Therefore four basis functions at orientations $\theta_k = k\pi/4,\ k = 0, \ldots, 3$ suffice to steer $A_{\theta,o}(\vartheta)$ approximately. In order to get four Hilbert pairs of basis functions, we have to supplement the even basis filter set by one additional filter. This expansion is redundant with respect to steerability but useful, if basis filter outputs are to be interpreted without steering.

The interpolation functions $b_k^\alpha(\theta)$ can be derived by applying the steerability eq. (6) to eq. (7).

Due to the properties of the $\cos^2$ function, an isotropic filter could be obtained simply by summing up basis filters for all orientations:

$$H_{\sigma,\mathrm{iso}}(\rho,\vartheta) = \sum_{k=0}^{3} \frac{1}{2} R_\sigma(\rho) \cdot A_{\theta_k,\alpha}(\vartheta) = R_\sigma(\rho) \tag{8}$$

The same principle yields isoscale filter within the scale interval $[\sigma_0, \sigma_2]$:

$$H_{\mathrm{iso},\theta,\alpha}(\rho,\vartheta) = \sum_{i=0}^{3} R_{\sigma_i}(\rho) \cdot A_{\theta,\alpha}(\vartheta) = A_{\theta,\alpha}(\vartheta) \tag{9}$$

2.3 Feature Integration

Once, the fixed step of convolution with the basis filters is completed, the obtained outputs $g_{k,\sigma,\theta,\alpha}(\boldsymbol{x})$ are linearly combined to obtain the saliency map $c(\boldsymbol{x})$:

$$c(\boldsymbol{x}) = \sum_{\ell} \sum_{k} \sum_{\sigma} \sum_{\alpha} w_\ell \, p_\sigma \, t_k^\alpha \, g_{\ell,\sigma,\theta_k,\alpha}(\boldsymbol{x}) \tag{10}$$

If the weights $w_\ell, t_k^\alpha, p_\sigma$ are set according to the specified target feature combination as listed in table 1, the resulting saliency map is supposed to have its highest local maxima at target locations.

colour			
$w_\ell,\ \ell =$	L	a	b
red	0	1	0
green	0	-1	0
yellow	0	0	1
blue	0	0	-1
bright	1	0	0
dark	-1	0	0

orientation/phase								
t_k^α $\alpha =$ $k =$	*even* 0	*even* 1	*even* 2	*even* 3	*odd* 0	*odd* 1	*odd* 2	*odd* 3
θ/*even*	$b_0^e(\theta)$	$b_1^e(\theta)$	$b_2^e(\theta)$	$b_3^e(\theta)$	0	0	0	0
θ/*odd*	0	0	0	0	$b_0^o(\theta)$	$b_1^o(\theta)$	$b_2^o(\theta)$	$b_3^o(\theta)$
NN	1/2	1/2	1/2	1/2	0	0	0	0

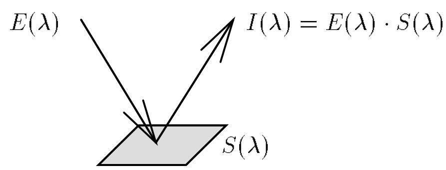
scale			
$p_i,\ i =$	0	1	1
0	1	0	0
1	0	1	0
2	0	0	1
NN	1	1	1

Table 1. Feature weights according to a target feature specification. The b_k^α are the interpolation functions.

3 Colour Constancy

3.1 Reflection and Sensing

First, we have to recall the physical process of reflection which mainly influences the light passing the lens of the camera. A single point on a Lambertian surface with spectral reflectance distribution $S(\lambda)$ receives light from an illuminant with spectral energy distribution $E(\lambda)$. The spectral distribution $I(\lambda)$ of the light reflected by the surface is obtained as the product



This process is referred to as *subtractive colour mixture*. After the spectral light $I(\lambda)$ has passed the lens, it hits three colour sensors with spectral sensitivities $Q_\ell(\lambda)$. The responses of the sensors, termed *tristimulus values*, are obtained by projecting the density $I(\lambda)$ to each of the sensitivity functions $Q_\ell(\lambda)$:

$$\begin{pmatrix} q_R \\ q_G \\ q_B \end{pmatrix} = \int_0^\infty I(\lambda) \begin{pmatrix} Q_R(\lambda) \\ Q_G(\lambda) \\ Q_B(\lambda) \end{pmatrix} \mathrm{d}\lambda \tag{11}$$

How do the values q_ℓ relate to surface reflectance? In order to answer this question, we first make the following assumptions[15, 12]:

1. The spectral reflectance distribution $S(\lambda)$ of the surface and the energy distribution $E(\lambda)$ of the illuminant is smooth and is well approximated by the weighted sum of spectral sensitivities $Q_\ell(\lambda)$:

$$E(\lambda) = \sum_{\ell \in \{R,G,B\}} e_\ell Q_\ell(\lambda); \qquad S(\lambda) = \sum_{\ell \in \{R,G,B\}} s_\ell Q_\ell(\lambda); \tag{12}$$

2. The sensor sensitivity functions are orthogonal i.e.

$$\int_0^\infty Q_k(\lambda)\, Q_\ell(\lambda)\, \mathrm{d}\lambda = \begin{cases} 1; & k = \ell \\ 0; & k \neq \ell \end{cases} \tag{13}$$

Then each sensor response $q_\ell; \ell \in \{L, a, b\}$ is obtained as a product of the illuminant value e_ℓ and the reflectance value s_ℓ:

$$\begin{pmatrix} q_R \\ q_G \\ q_B \end{pmatrix} = \begin{pmatrix} e_R \cdot s_R \\ e_G \cdot s_G \\ e_B \cdot s_B \end{pmatrix} \tag{14}$$

By this, the potentially infinite dimensional illuminant and surface reflectance functions are represented jointly in a three dimensional space spanned by the sensitivity functions $Q_\ell(\lambda)$. This is a reasonable assumption for most natural surfaces and illuminants. In fact, the principle component analysis applied to a large set of natural surfaces which Maloney[12] performed, roughly yields the spectral sensitivities of retinal photoreceptors as the most important three eigenfunctions.

The second assumption mentioned above, the orthogonality of the $Q_\ell(\lambda)$, is a stronger restriction of the model.

3.2 Properties of Lab–Colour Space

Colour constancy is equivalent to get rid of the factors e_ℓ of eq. (14) which come from illumination. Now we transform the sensor responses q_ℓ to the Lab–colour space. Due to its resemblance with the logarithmic function, the cubic root function in eq. (1) approximately maps products to sums, and factors to summands. Consequently, the Lab–coordinates q_ℓ of the reflected light are obtained by:

$$\begin{pmatrix} q_L \\ q_a \\ q_b \end{pmatrix} = \begin{pmatrix} e_L \\ e_a \\ e_b \end{pmatrix} + \begin{pmatrix} s_L \\ s_a \\ s_b \end{pmatrix} \tag{15}$$

where e_j and s_j are the *Lab*–coordinates of e_ℓ and s_ℓ, respectively.

As in an earlier colour vision model by Frei and Baxter[5], we could have used the logarithmic characteristics, but we preferred the *Lab*–space because of its uniform colour distance measure. We will exploit this property for attention and coding in future work.

3.3 Linear Filtering

The *subtractive colour mixture*, we observe in the physical world has turned out to be an *additive colour mixture* in the perceptual world.

The task is now to extract the summands s_ℓ from eq. (15). If we assume that spatial variations in illumination are slower than those in reflectance[15], the problem is solved with a linear filter. Looking then at the spatial frequency domain, the low frequency components come from the slowly varying illuminant, while the high frequency components are due to changes in surface reflectance. By means of band–pass Andersson filters, described in section 2, the low frequency components are suppressed. The filter outputs only carry the reflectance component $s_\ell(\boldsymbol{x})$ of the colour signal.

4 Experimental Results

We tested our algorithm in the framework of a visual attention system, that was inspired by Ahmad's[1] VISIT. The purpose of VISIT was to simulate some psychological experiments upon human visual search. A typical test pattern in these experiments contains coloured rectangles on a homogenouos background. The rectangles are either in vertical or in horizontal orientation. A visual search task is defined uniquely by colour and orientation of the search target.

The aim is to detect the target by means of the output of an oriented Andersson filter[2] on *Lab*–values which is matched to the orientation, scale and colour of the target. The more powerful scheme of multiscale multiorientation analysis here degrades to a template matching. Colour invariant detection is achieved, if the filter responses at the center of each object is independent of the illuminant colour. To test also the physical part of our model, we did not simulate the coloured illumination on the computer, but realized it with a colour filter in front of the spot lights.

Figure 2 shows the a component of the original image under white (2a) and red illumination (2b). In figure 2c) the filter response to the a component of the white illuminated image is depicted. We have left out the response to the red illuminated image, because it is indistinguishable from the white illumination response, when printed on paper.

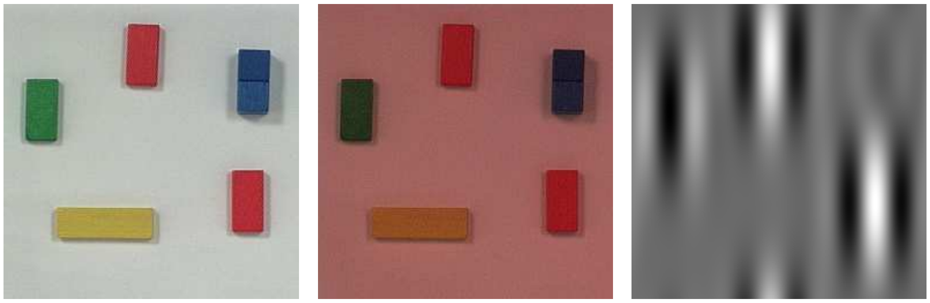
5 Summary

We presented a multiorientation multiscale representation for colour image signals which is largely invariant with respect to intensity and colour of illumination.

Many computer vision problems have been done without colour information. Therefore colour processing issues are often neglected or considered as a byproduct of luminance processing. In attentive architectures however, colour processing is essential, because it is one of the most important features in preattentive vision. Since attention is mostly attracted by reflecting objects, not by illumination, preattentive features have to work colour and lightness constant. On the other hand there are many retinex–like colour constancy algorithms which involve isotropic filters and thus discard orientation. In our approach, both colour constancy and lightness constancy is achieved by the same unifying steerable filter approach while preserving the information of local orientation.

References

1. S. Ahmad. *VISIT: An Efficient Computational Model of Human Visual Attention.* Phd thesis, University of California, Berkeley, 1991.



a) b) c)

Fig. 2. The–a component p_a of a white (a) and red (b) illuminated camera sensed test image and the corresponding filter response (c). Red illumination (b) globally yields higher a–values, but almost the same filter response. No normalization was applied to the images. The responses at the borders of figures (c) are due to the cyclic effect of frequency domain filtering

2. M.T. Andersson. *Controllable multidimensional filters and models in low level computer vision.* PhD thesis, Linköping University, 1992.
3. J. D. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two–dimensional visual cortical filters. *Opt. Soc. Am. A*, 2(7):1160–1169, 1985.
4. W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Trans. PAMI*, 13(9):891–906, 1991.
5. W. Frei and B. Baxter. Rate distortion coding simulation for color images. *IEEE Trans. Com.*, 25:1349–1384, 1977.
6. B. V. Funt and G. D. Finlayson. Color constant color indexing. *IEEE Trans. PAMI*, 17(5):522–529, 1995.
7. G.-J. Giefing, H. Janßen, and H. Mallot. Saccadic object recognition with an active vision system. In John Wiley & Sons, editor, *10th European Conference on Artificial Intelligence, Vienna*, pages 803–805, 1992.
8. R. Herpers, H. Kattner, H. Rodax, and G. Sommer. GAZE: An attentive processing strategy to detect and analyze the prominent facial regions. In *International workshop on face– and gesture recognition, IWAFGR95, Zurich*, June 1995.
9. B. K. P. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, pages 277–299, 1974.
10. C. Koch and S. Ullman. Shifts in selective attention: towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227, 1985.
11. E. H. Land. The retinex. *Am. Scientist*, 52, 1964.
12. L. T. Maloney and B. A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *Opt. Soc. Am. A*, 3:29–33, 1986.
13. A. M. Treisman and G. Gelade. A feature integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.
14. J. K. Tsotsos. Analyzing vision at the complexity level. *The Behavioral Brain Sciences*, 13:423–469, 1990.
15. B. A. Wandell. The synthesis and analysis of color images. *IEEE Trans. PAMI*, 9:2–13, 1987.

This article was processed using the LaTeX macro package with LLNCS style