Fundamental Restrictions of the Trichromatic Imaging

Vitali V. Gavrik
Cologne, Germany

Abstract

The standard color matching functions show that a trichromatic imaging system could not in principle render nearly a half of colors (especially among the saturated ones) seen by a normal observer even if the primaries used were monochromatic spectral colors of maximum saturation. The trichromatic systems cannot also accurately render the rest of colors since they are unable to simulate the non-existing linear transforms of the alternating matching functions of a triad of real primaries to a triad of non-negative spectral sensitivities of the three hypothetical cones. Even in its modernized, opponent-color form, the trichromatic hypothesis cannot obviate numerous experimental inconsistencies and is forced to involve a kind of chromatic influence of the retinal rods. The recently proposed approach to consider the rod and the cone as two color receptors that intrareceptorally perform the blue-yellow and red-green opponent color separation, respectively, eliminates the controversy. It permits to determine and simulate (a) the four necessary chromatic imaging primaries corresponding to the four basic visual sensations and (b) the spectral sensitivities of the respective pairs of differential opponent receptors within the rod and cone, as well as (c) the spectral location of colors which evoke the primary sensations in their most intensive and saturated states. The approach appears to provide an actual separation of chromatic and achromatic (white) components of a color in the form of their orthogonal (non-correlated over the visible spectrum) matching factors.

Introduction

The color rendering in the state-of-art trichromatic imaging run across some unavoidable restrictions,[1-4] contributing to the position by W.D. Wright: "How Thomas Young would have smiled if he could realize the endless controversy to which his suggestion would lead!".[5] Th. Young's predecessors usually did not take into account the saturation when reported that every color could be rendered with a mixture of three primaries.[6]

Over a half of the CIE XYZ chromaticity diagram (Fig. 1) cannot be rendered by the sums (positive linear combinations) of the three standardized colorimetric R, G, and B primaries (the monochromatic 700, 546.1, and 435.8-nm radiations, respectively). The colors need a 'minus-matching', unavoidable at any choice of three real primaries. A primary should be 'subtracted' that is added to the color itself as the standard colorimetric procedure prescribes. That strongly desaturates such a color and is useless in the imaging of natural scenes even if it would have been possible differently to add the primaries to the various scene colors themselves. Even if a display had disposed of the maximum-saturation spectral primaries above, it would have been unable to render the chromaticities of subject colors that lay out of the standard RGB triangle shaded on Fig. 1.

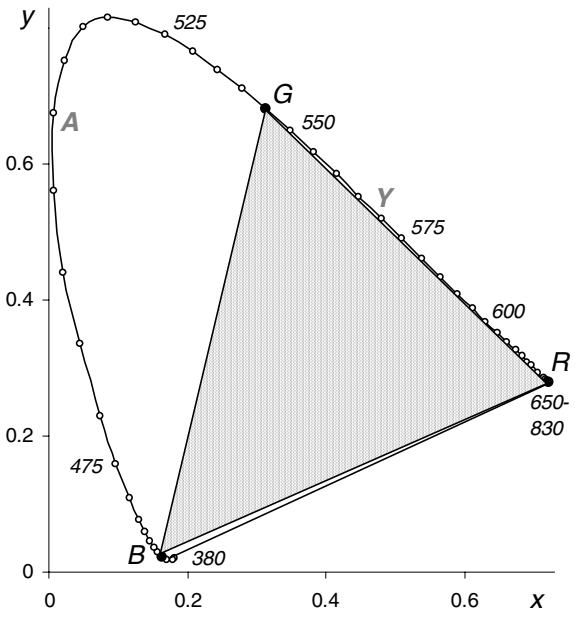

Figure 1. The RGB triangle of colors only reproducible as the weighted sums of the standard colorimetric primaries on the CIE 1961 chromaticity diagram.

Th. Young did not dispose of the data as well as other evidence against his hypothesis of the three kinds of independent chromatic receptors within the retina. Till present, there is no direct experimental evidence for the different spectral classes of photoreceptor cells but the anatomical and photoelectric differences between the rod and cone.[7] Nothing but the two stable visual pigments have been detected in the living retinas by the method of comparison of their spectral reflection after long dark adaptation and after strong illumination: rhodopsin (max. 500 nm) in the rods that predominate in the retinal periphery,[8] and iodopsin (max. 560 nm) in the cones that predominate in the fovea.[9] The same stable visual pigments only were chemically separated from the retinas of ground vertebrates.[7] Direct photoelectric measurements

of the light-sensitive outer segments of human rod[10] and cone[11] reliably detected only the two light-sensitive pigments above at moderate illumination.

Their labile photochemical intermediates were frequently detected after partial bleaching of living retinas.[12-14] In dead retinas that could not have properly recovered after a strong surgical illumination, the microspectrophotometry of outer segments[15,16] not seldom demonstrated even a predominance of the long-living intermediates[12,17] usually described as the 'M-cone pigments.' Even the late indirect estimations of the 'ratios of different cones' (for instance, see the papers on 'Chromatic topography of the retina', *J. Opt. Soc. Am.*, v. A17, #3, 2000) have varied several times at different light exposures and even by orders of magnitude if different trichromatic assumptions were used for the calculations. Thus, the trichromatic hypothesis goes on with demonstrating its general inconsistency with the basic experimental data.

The imaginary X, Y, and Z 'primary colors' were chosen by Guild and Wright from the infinite diversity of equivalent linear transformations of the color matching functions to provide the standard color description with positive numbers only. It is sometimes thought that the positive matching functions of the irreal colors are the closest to the spectral sensitivities of the three cones. E. Schrödinger mathematically showed[18] that the color matching data could alternatively correspond to the two independent alternating opponent chromatic axes predicted by E. Hering. The color as such remained a *two-dimensional* feature related to the two chromatic opponent axes and its luminosity was assumed to correspond to an extra achromatic axis. Two basic visual sensations on an opponent axis were complementary: they never produced a mixed blue-yellow or red-azure hue and reduced each other to a white. A pair of basic sensations from the two axes produced mixed blue-green, blue-red, yellow-green, or yellow-red sensations.

The opponent electric responses along the visible spectrum were directly recorded by G. Svaetichin from individual photoreceptor cells.[19,20] Nevertheless, the responses were treated as a result of a further neural processing of signals from the three cones after they had been found in various retinal structures along the neural pathways.[21] The holy trinity have enjoyed a hot believe of its confession even if the multiple experimental inconsistencies implied not three but only two independent receptors to suffice.[22,23] The blue/yellow or red/blue-green separation has been recently shown optically to occur in a manner similar to a plus-minus signal difference between two portions of light-sensitive agent separated by a spectral filter along the light path. The brightness could be a sum of partial signals.[22-24]

Physiological Mechanism of Intra-Receptor Color Opponency

An adequate physiological mechanism was proposed for the cone to perform the blue-green/red chromatic separation on the base of iodopsin ('L-cone' pigment) and the rod to perform the blue/yellow separation on the base of rhodopsin. Their long-living colored photochemical intermediates, such as metarhodopsin III (460 nm, half-decay of 90 s) and so called M-cone pigment (530 nm)[7,17] have to act as optical filters, differently modifying the spectral sensitivities in the two anatomical parts of outer segment that oppositely electrically contacted the photoreceptor cell membrane.[25,26] The simulated spectral sensitivities of the receptor pairs in the rod and cone were similar to those calculated by Schrödinger from color matching data and to those of dichromatic color deficiencies.[17,23] It has been shown that the opponencies occur after a photopic illumination accumulates enough the colored photochemical intermediates within an outer segment. Their relative spectral properties did not practically vary with orders of magnitude photopic illumination increase.

The chromatic states of the rod and cone have been shown achievable if the axial optical density of visual pigment was over 1.3 and over 0.7 for the rhodopsin and iodopsin, respectively. The estimated densities were usually lower and widely varied because of lacking in the reference to the actual photoreceptor number per unit area within the measured retinal regions, as well as because of obvious partial recovery of bleached visual pigments and the disregarding for the photopigment inhomogeneities in the lamellar structure of OS. The estimates have been reconsidered[27] with using the latest data on the reflection density differences between dark adapted and 'fully bleached' states of retina, on the photoreceptor mosaic, and the actual diameters of outer segments in the areas where there were practically either no rods or no cones.

The late reflectometry data on living human retinas[8,9,28,29] that can be related to the data on the geometry of the photoreceptor mosaic in the measured retinal areas[30,31] gave 0.90 ± 0.15 in the cone (at the single ~560-nm spectral absorption maximum of iodopsin in the rod-free central fovea) and 1.7 ± 0.6 in the rod (at the single ~500-nm spectral absorption maximum of rhodopsin in a peripheral retinal zone with the largest number of rods per unit area).[27]

Using the transversal microspectrophotometry of the outer segment of photoreceptor cells also led to a strong underestimation of the values since the specific absorbance of photopigment was commonly calculated without taking into account the typical optical inhomogeneities in photopigment distribution.[15] The grating of photopigmented lamellas strongly decreased the measured optical density of the outer segment similarly to other systems[32] with optical inhomogeneities of such a small size.[25,26] The improved estimates of the axial absorbance were 0.75 ± 0.15 in the cone and 1.55 ± 0.25 in the rod, confirming the reflectometric estimates above within the error.[27]

Another crucial condition for the opponent chromatic states to possess their spectral properties close to those of the human dichromasies is that the length of the basal part should be about 1:20 and 1:2.8 of the length of the entire outer segment in the rod and cone, respectively. After our calculations had been performed, the direct indications have been found in the physiological literature that the basal part takes actually "about a twentieth" in the rod outer segment and "a third" in the cone outer segment.[25,26] Under all the conditions, the relative spectral properties of the opponencies remained practically unchanged with orders of magnitude variations in the photopic illumination. They did not depend on the length and diameter of the outer segments that are known

significantly to vary with the retinal location of photoreceptor cells.

The tetrachromatic model provided for simple physical explanation the numerous experimental inconsistencies of trichromatic approach.[7,17] For example, it suggested the long-wave limit of the yellow primary sensation to lie always at least by 30 nm lower than that of the red primary sensation. Various experimental techniques have shown that even in the persons with the normal color vision and the deuteranopes ('L-cone dichromates'), the long-wave limit of the red sensation really overcomes by 30-40 nm the limit for the yellow sensation.[19,20,33,34] That strictly contradicts to one of the basic trichromatic assumptions that the yellow sensation could have resulted from the L-cone response summarized with the M-cone response. Any attempts to calculate the hypothetical three cones whose signals would have been processed in an opponent manner by some after-receptor retinal structures must but did not explain the spectral difference.[35]

Orthogonality of the Opponent Chromatic and Achromatic Matching Functions

For the general colorimetry purposes, any color can be matched with two real non-opponent primaries only, say, the red and blue, instead of the usual three. A similar principle underlies so called hue-cancellation techniques. With adding or 'subtracting' the red primary, one can control the content of the red and blue-green. With adding or subtracting the yellow primary, the content of the yellow and complementary blue can be controlled. With adding or subtracting both the primaries simultaneously, one can control the contribution of white.

Even if it could be successfully applied to the tasks of single-color measurement and characterization, the desaturation techniques cannot be used for color imaging purposes. The colors in a trichromatic image are in general less saturated than in a scene and that cannot significantly improved without involving a fourth primary.[1] The four imaging primaries could directly represent all the basic chromatic sensations of vision that are seen by the normal observers with no admixture of another primary in the neutral zones of the red-azure and blue-yellow opponencies that should be close to those of the tritanopic (about 570 and 450 nm) and protanopic or deuteranopic chromatic axes (about 500 and over 650 nm). Such an approach appears to provide for a simplest adjustment of four primaries with respect to the gray imaging scale,[7,17,24]

Adequately to render mixed primary sensations in a natural image requires the spectral ranges of imaging primaries to simulate the sensitivity ranges of respective opponent pairs of chromatic receptors in the rod and cone (Fig. 2). Such curves were first obtained by Schrödinger,[18] then by Judd & Wyszecki,[37] and Jameson & Hurvich[38] as equivalent transfomations of color matching data. Similar dependencies with the same neutral zones were also recently calculated from the spectral absorption of rhodopsin in the rods and iodopsin in the cones at a sufficiently high equilibrium concentration of their long-living photochemical intermediates accumulated at photopic illumination.[22,23]

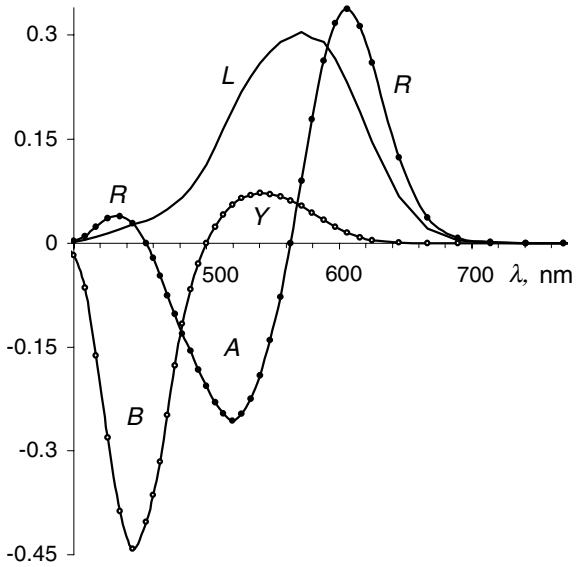

Figure 2. Normalized opponent matching functions: blue-yellow (1), red-azure (2), and achromatic (3). The latter is taken "from subjects' own direct comparison of luminosity factors" of the colorimetric primaries used (the last column in Table 4[36]).

The current trichromatic imaging systems were recently found[7,17] to be practically adjusted by the trial-and-error method so that they roughly simulated the main spectral sensitivity regions of partial vision receptors and delimit them near to the principal 495- and 570-nm neutral points of the blue-yellow and red-azure opponencies of human vision. A complete simulation of the spectral sensitivities of the partial chromatic vision channels was in principle unachievable without using two independent middle-wave imaging channels (yellow and azure) instead of the usual yellow-green hybrid.[24]

Nevertheless, the four imaging primaries, even if they were monochromatic spectral colors accurately corresponding to the neutral points of the opponencies, would inherit some shortcomings of current color imaging. Fig. 2 shows that the primaries would lie out of the spectral zones of the maximum chromatic intensity of the respective basic sensations of vision.[18,23] For instance, the azure basic sensation is seen without another chromatic admixture around 500 nm but has its maximum intensity around 520 nm. Such a 500 nm primary of lowered saturation cannot impart the necessary chromatic intensity when used to render a subject color from the spectral range of higher intensity of the basic sensation. It seems obvious that the actual spectral saturation of all the four basic color sensations could be estimated after the chromatic opponent functions above (a) would be normalized so that to their sum equated to the sum of the original RGB matching functions,[39] and (b) related to a function independently describing the achromatic luminance signal.

It seems to be commonly expected that the photoreceptor process functions which really underlie the color separation, should (a) be sufficiently insensitive of slight variations in the transformation conditions, and (b)

possess some mathematical properties that would imply a kind of the mutual independence of receptor responses over the visible spectrum. The second condition would provide the highest color coding efficiency of the spectral properties of various objects in the natural scenes.

The opponent chromaticity functions of Fig. 2 have been calculated from the RGB-matching data that had been published by Stiles and Burch[40] for the 10° observation field and underlain the CIE 1961 additional colorimetric observer. The data are considered in the literature as uniquely published without some usual 'questionable' transformations.[35]

An Independent Components Analysis techniques[41] has been applied that was previously shown to permit detection of some functional properties of the actual physical basis ('process functions') in a linear structure of multivariate data spread. It is applicable to either complete group of experimental functions that, within an error, are actually nothing else as the linear combinations of several 'process functions' with random weighting factors. It has been shown that all the three process functions are proportional to each other or at least one of them is zero after 660-680 nm. Whereas there are all the reasons to assume that the RA and L functions are really proportional to each other even also at shorter wavelengths, the BY chromaticity function should be zero after the long-wave limit of the yellow sensation.[42] Also in the two spectral ranges, about 455-465 and 570-590 nm, all the three process curves should be close to linear functions, say have overlapping parts of their near-to-inflection regions.

The main conditions to perform their transformation were the usually used neutral zones where the opponencies produce the zero chromatic signal. Since the actual neutral zones of the opponencies in the normal color vision can differ from those of the defective dichromatic vision and are not known with sufficient accuracy, we used roughly to equate them to the abscissas of the original work with no interpolation.[40]

The linear transformation to the RA curve has been based on the neutral zones of the red-azure chromatic opponency that were directly seen by a tritanope around 570 and 460 nm. The estimates of the second neutral zone were not certain and varied from 410 to 470 nm in the literature. The calculated curves held their general shape with extensive variation in the assumed spectral position of the zero-chromaticity points. The linear transformation of the color matching data to the BY curve was based on the neutral zones about 500 and over 660 nm of the blue-yellow chromatic opponency that are directly seen by a deuteranope (a protanope sees the second zone as a black one). The BY curve practically did not vary while the second zero-chromaticity zone was being represented by any experimental point after 680 nm.

Along with the general shape stability, the chromatic curves turned out to possess another important property. They were orthogonal to each other within the error. The scalar products of respective unit vectors, such as those on Fig.2, differed from zero within ± 0.05. The orthogonality of chromatic functions held when the spectral position of the neutral zones has been varying from 494 to 507 nm (BY), 444 to 465 nm, and 563 to 580 nm (RG) - the original abscissas used by Stiles & Burch.[36,40]

The achromatic function L has been orthogonal to both the chromatic opponent functions but within a more restricted spectral range for the latter neutral zone (close between 571 and 565 nm). That might result from the known uncertainty of what a kind of luminosity function is really representing the achromatic, white contribution to the spectral colors.[35] Another possible ground for the spectral shift could be that our calculations corresponded to an imaginary light source with the equal distribution of visible radiation energy over the spectrum. At a typical color temperature of natural illumination, the radiations of various wavelengths should perceptibly differ in their contributions to the orthogonality measures above.

Thus, from all the possible length ratios of partial opponent receptors within the outer segment of the rod and cone, the evolution has chosen those producing the three non-correlated responses over the visible spectrum: two opponent chromatic and one achromatic. Unfortunately, there were no sufficient number of transformation conditions available for independently to calculate the L-curve. The achromatic axis of the color space was assumed to be better characterized by the 10° luminosity estimate proposed by Stiles & Burch.[36] The function appears definitely to need its additional investigation.[35]

Nevertheless, the result implies that relating the two independent chromatic response functions to the independent achromatic response function could give us two opponent spectral saturation curves: one for the red and blue-green basic sensations, and the second for the blue and yellow basic sensations of vision. In an imaging context, some of the irreproducible chromaticities out of the RAYB quadrangle (Fig. 2) may imperceptibly differ from the reproducible chromaticities at its edges. The maximum-saturated red primary seems easily to be obtained in its pure state, starting from 650 nm. The saturation of blue primary in its unmixed chromatic state at 460 nm is also close to its maximum saturation at 440 nm. Otherwise, it would be reasonable to choose an imaging primary at the wavelength of maximum saturation of the respective basic sensation of vision and then to perform a proper linear transformation in the imaging system. Say, the 510 nm primary may be chosen for the first middle-wave imaging channel and 540 nm primary for the second where the respective blue-green and yellow vision primaries possess their maximum saturation.

References

1. H. Heki, Design of color reproduction required for portrait photography, *Intl. Cong. Imag. Sci*., IS&T, v.1, 212-213 (1998).
2. Fuji Photo Film, *Fujicolor Superia 200. Data Sheet* (1998).
3. B. Hill, Multispectral color technology: A way towards high definition color image scanning and encoding, *Proc. SPIE,* v. 3409, 2-13 (1998).
4. F. König & P.G. Herzog, On the limitations of metameric imaging, *Proc. PICS*, IS&T, 163 - 168 (1999).
5. W.D. Wright, The present status of the trichromatic theory, *Docum. Ophthal.,* v. 3, 10-23 (1949).

6. P.K. Kaiser & R.M. Boynton, *Human Color Vision*, OSA, 1996.
7. V.V. Gavrik, Tetrachromacy of Human Vision: Spectral Channels and Primary Colors, *Proc. SPIE,* v. 4421 *(9th Cong. Intl. Colour Association),* AIC, 4 p. (2001).
8. H. Ripps & A.G. Snapper, Computer analysis of photochemical changes in the human retina. *Computers in Biology and Medicine,* v. 4, 107-122 (1974).
9. P.E. Kilbride, J.S. Read, G.A. Fishman & M. Fishman, Determination of human cone pigment density difference spectra in spatially resolved regions of the fovea, *Vision Res*., v. 23, 1341-1350 (1983).
10. D.A. Baylor, T.D. Lamb & K.-W. Yau, The membrane current of single rod outer segments, *J. Physiol.,* v. 288, 589-611 (1979).
11. J.L. Schnapf, T.W. Kraft & D.A. Baylor, Spectral sensitivity of human cone photoreceptors, *Nature,* v. 325, 439-441 (1987).
12. P.K. Brown & G.Wald, Visual pigments in human and monkey retinas, *Nature,* v. 200, 37-43 (1963).
13. H. Ripps & R.A. Weale, Photo-labile changes and the directional sensitivity of the human fovea, *J. Physiol.,* v.173, 57-64 (1964).
14. A. Roorda & D.R. Williams, The arrangement of the three cone classes in the living human eye, *Nature,* v. 397, 520-522 (1999).
15. H.J.A. Dartnall, J.K.Bowmaker & J.D.Mollon, Human visual pigments, *Proc. Roy. Soc. Lond.,* v. B220, 115-130 (1983).
16. E.F. MacNichol, J.S. Levine, R.J.W. Mansfield, L.E. Lipetz & B.A. Collins, Microspectrophotometry of visual pigments in primate photoreceptors, In: *Colour Vision*, J.D. Mollon, ed., 13-38 (1983).
17. V.V. Gavrik, Four Spectral Channels for Natural Imaging of Scene Colors, *Proc. 16th Intl. Conf. on Digital Printing Technologies*, IS&T, 560-565 (2000).
18. E. Schrödinger, Über das Verhältnis der Vierfarben- zur Dreifarbentheorie. *Sitzungsber. Akad. Wiss., Wien*, Bd. IIa, Nr. 471-490 (1925) [*Farbe*, v. 41, # 4-6, 178-197 (1995)].
19. G. Svaetichin, Spectral response curves from single cones. *Acta Physiol. Scand.,* v. 39, Sup. 134, 17-46 (1956).
20. G. Svaetichin & E.F. MacNichol, Retinal mechanisms for chromatic and achromatic vision, *Ann. N.Y. Acad. Sci.,* v.74, 385-404 (1958).
21. F. Ratliff, Mach Bands: Quantitative Studies on Neuronal Networks in the Retina. San Francisco, 1965.
22. V.V. Gavrik, Opponent rhodopsin receptor in rod and opponent iodopsin receptor in cone underlie the human colour perception, *Perception*, v. 27S (ECVP98, Oxford), 171 (1998).
23. V.V. Gavrik, A mechanism for single-pigment color opponency in a photoreceptor cell, *Proc. Intl. Joint Conf. on Neural Networks*, IEEE, 174-177 (1999).
24. V.V. Gavrik, The independent yellow channel for tetrachromatic imaging. *Proc. NIP17: Intl. Conf. on Digital Printing Technol.,* IS&T, 462-464 (2001).
25. A.I. Cohen, The fine structure of the extrafoveal receptors of the Rhesus monkey, *Exp.Eye Res*., v.1, 128-136 (1961); Rods and cones. In: *Physiology of Photoreceptor Organs*. M.G.F. Fuortes, ed., Springer, Berl., 63-110 (1972).
26. R.H. Steinberg, S.K. Fisher & D.H. Anderson, Disc morphogenesis in vertebrate photoreceptors, *J. Comp. Neurol*., v. 190, 501-518 (1980).
27. V.V. Gavrik Axial absorbance of outer segments from reflectometric and microphotometric data. *Optics Express 9*, *#8* (*OSA & UCI Color Workshop, Irvine, CA*), PC8 (2001).
28. H. Ripps & R.A. Weale, Flash bleaching of rhodopsin in the human retina. *J.Physiol*., v. 200, 151-159 (1969).
29. T.T.J.M. Berendschot, J. van de Kraats & D. van Norren, Foveal mosaic and visual pigment density in dichromats. *J. Physiol*., v. 492, 307-314 (1996).
30. M. Curcio, A. Christine & A.E. Hendrickson, Organization and development of the primate photoreceptor mosaik. *Prog. Ret. Eye Res.*, v. 10, 89-120 (1991).
31. C.A. Curcio, K.R. Sloan, O. Packer, A. Hendrickson & R.E. Kalina, Distribution of cones in human and monkey retina: Individual variability and radial asymmetry. *Science*, v. 236, 579-582 (1987).
32. V.V. Gavrik, The maximum covering power of light-absorbing substances under imaging conditions. *Imag. Sci. J.*, v. 44, # 1, 5-7 (1996).
33. B.D. Drujan & M. Laufer, eds., *The S-potential*, Alan R. Liss, Inc., NY, 324 p., 1982.
34. J. Toyoda, M. Murakami, A. Kaneko & T. Saito, eds. *The Retinal Basis of Vision,* Elsevier, Amst., 290 p., 1999.
35. A. Stockman & L.T. Sharpe, Cone spectral sensitivities and color matching. In: K. Gegenfurtner & L.T. Sharpe, eds. *Color vision: from genes to perception* Cambr.Univ. Press, 51-85, 1999.
36. W.S. Stiles & J.M. Burch, N.P.L. colour-matching investigation (1955). Mean reasults for pilot group of ten subjects. *Opt. Acta*, v. 2, 176-181 (1955).
37. D.B. Judd & G. Wyszecki, *Color in Business, Science and Industry.* 3rd ed., Wiley, NY, 1975.
38. L.M. Hurvich, *Color vision*. Sinauer, Sunderland, MA, p. 328, 1981.
39. W.S. Stiles & J.M. Burch, N.P.L. colour-matching investigation: Final report (1958). *Opt.Acta*, v. 6, 1-26 (1959).
40. V.V. Gavrik, Criteria for linear structure of over-error variance of experimental dependencies and detection of physical basis from principal components. *Proc. 3rd Intl. Conf. 'Independent Component Analysis and Blind Signal Separation',* UCSD, 200-205 (2001).
41. V.V. Gavrik (2001). The yellow sensation: Spectral independence of the L-cone. *Optics Express*, v. 9, # 8 (*OSA & UCI Color Workshop, Irvine, CA*), PC9.

Biography

Vitali V. Gavrik received his Ph.D. degree in Imaging Science and Engineering (1974) and his D.Sc. degree in Physics and Mathematics (1996) from the Federal Research Center 'S.I. Vavilov State Optics Institute', St.Petersburg, Russia. Since 1965 to 1996, he worked at the research laboratories of the FRC. His work has primarily focused on the quality metrics of color images, color perception, photophysics and photochemistry of imaging processes. He is a member of the IS&T and the Russian Optical Society.