Not so Intelligent Multimedia Systems

WE DON'T NEED NO EDUCATION!

Tuesday 18-12-'12, 15:00-18:00, SP C1.110

It is **not** open notes. You can bring a calculator, though, and a pen, a pencil!. Alcohol?

Exercises I

EXERCISE 1:

To calculate the color of light sources, the following intuitive color models are used: intensity I, chromaticity xy, hue H and saturation S. Let's assume, for simplicity reasons, that sunlight S is given by X = Y = Z = 100. Further, let X = 100, Y = 100 en Z = 150 be the values for a given artificial lamp A.

(a) Calculate the intensity I of the two light sources S and A.

Intensity = (R+G+B)/3 or (X+Y+Z)/3

$$I_S = 100$$

$$I_A = 116.6$$

(b) Calculate the chromaticity values x = X/(X + Y + Z), y = Y /(X + Y + Z) and plot these in the chromaticity diagram given in Figure 1.

x = R / (R +G +B)

So,

$$x_S = \frac{100}{100+100+100} = \frac{1}{3} = 0.333$$

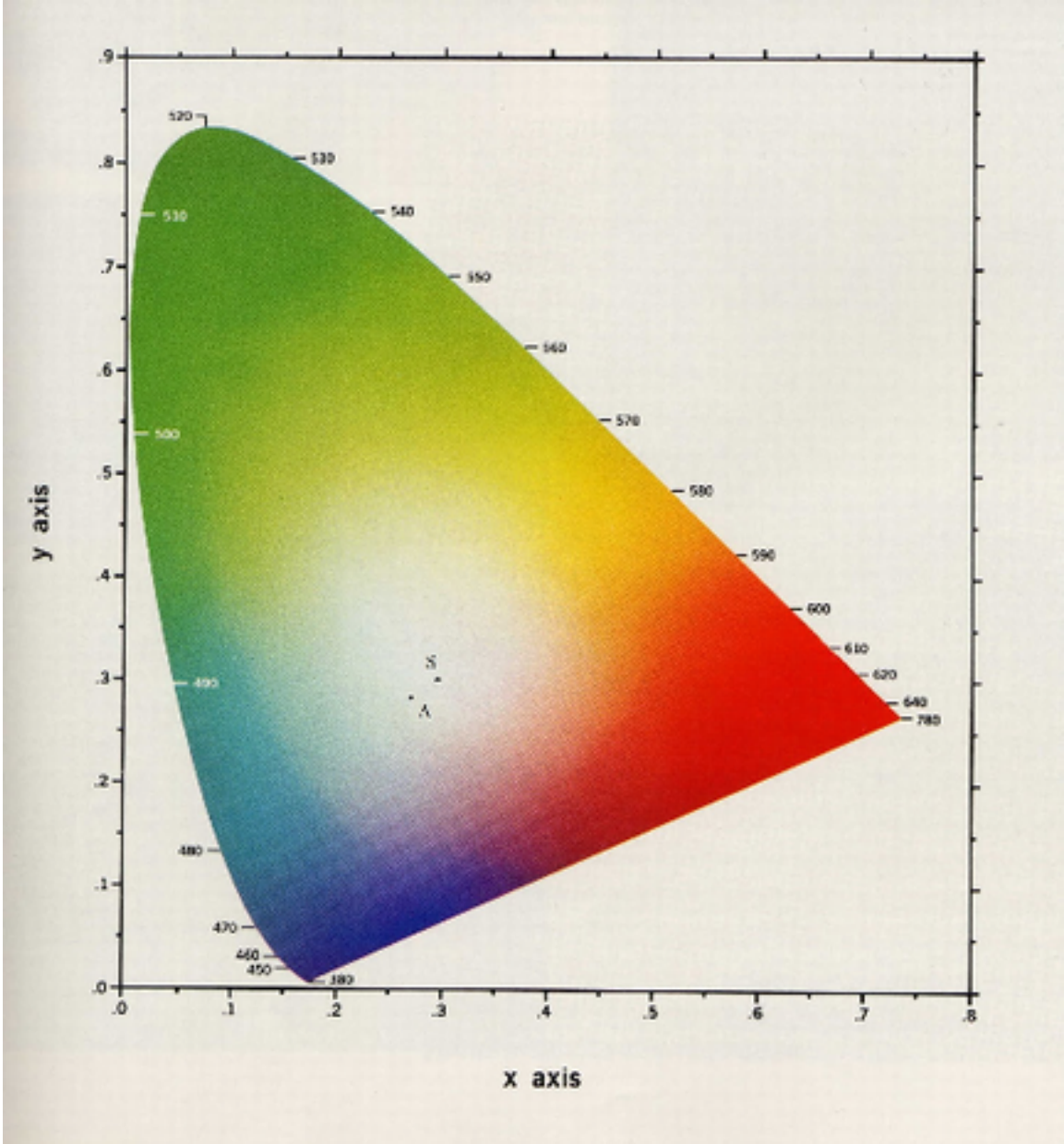
$$y_S = \frac{100}{100+100+100} = \frac{1}{3} = 0.333$$

$$z_S = \frac{100}{100+100+100} = \frac{1}{3} = 0.333$$

$$x_A = \frac{100}{100+100+150} = \frac{2}{7} = 0.286$$

$$y_A = \frac{100}{100+100+150} = \frac{2}{7} = 0.286$$

$$z_A = \frac{150}{100+100+150} = \frac{3}{7} = 0.428$$



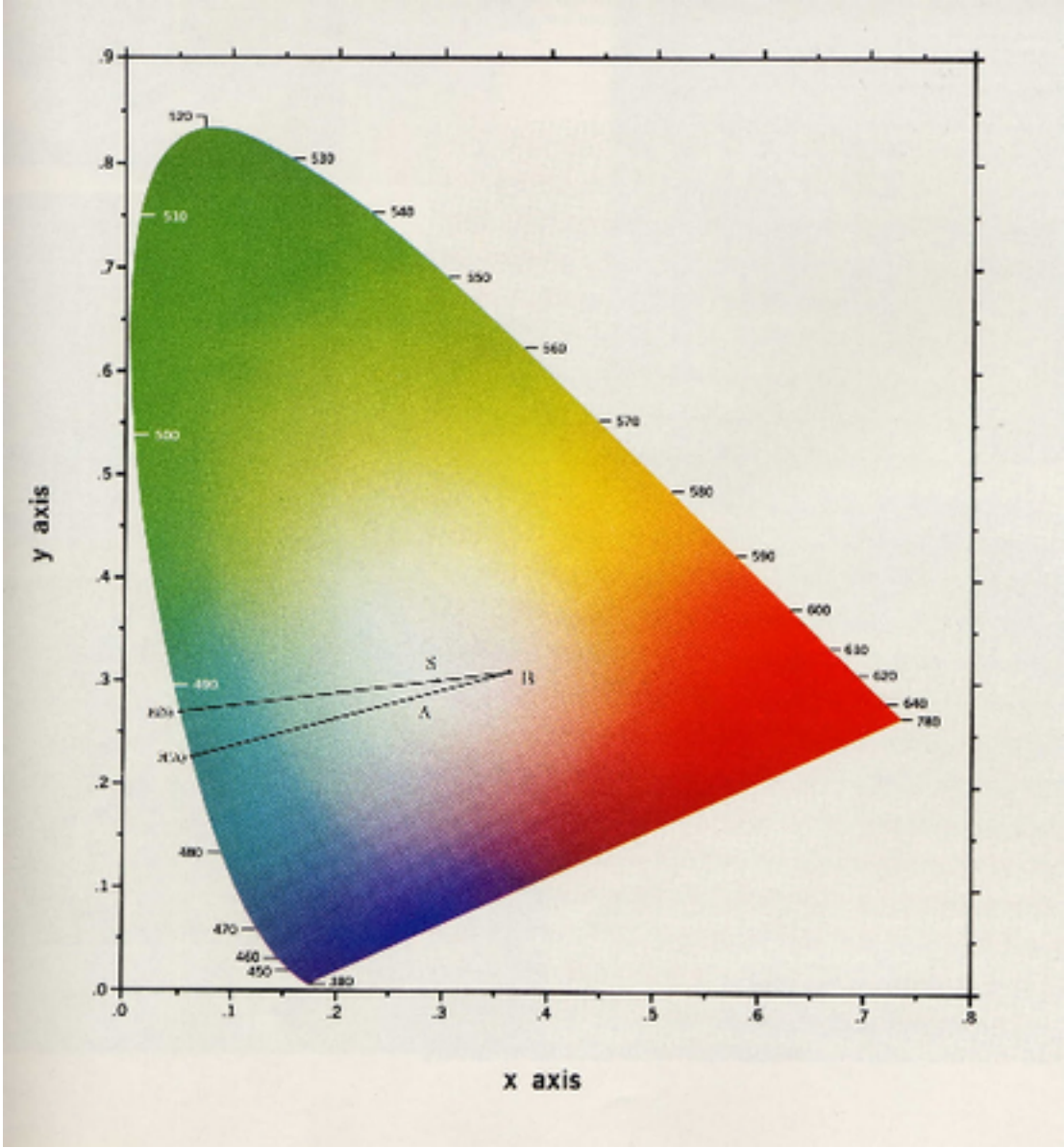
(c) What the estimated hue of the light sources S and A with reference white light B at X = 120, Y = 100 and Z = 100.

$$x_B = \frac{120}{120+100+100} = 0.375$$

$$y_B = \frac{100}{120+100+100} = 0.312$$

$$z_B = \frac{100}{120+100+100} = 0.312$$

You get the estimation of the hue by drawing a line from the reference white light B through S and A.
Hue of S $\approx$505 nm
Hue of A $\approx$480 nm

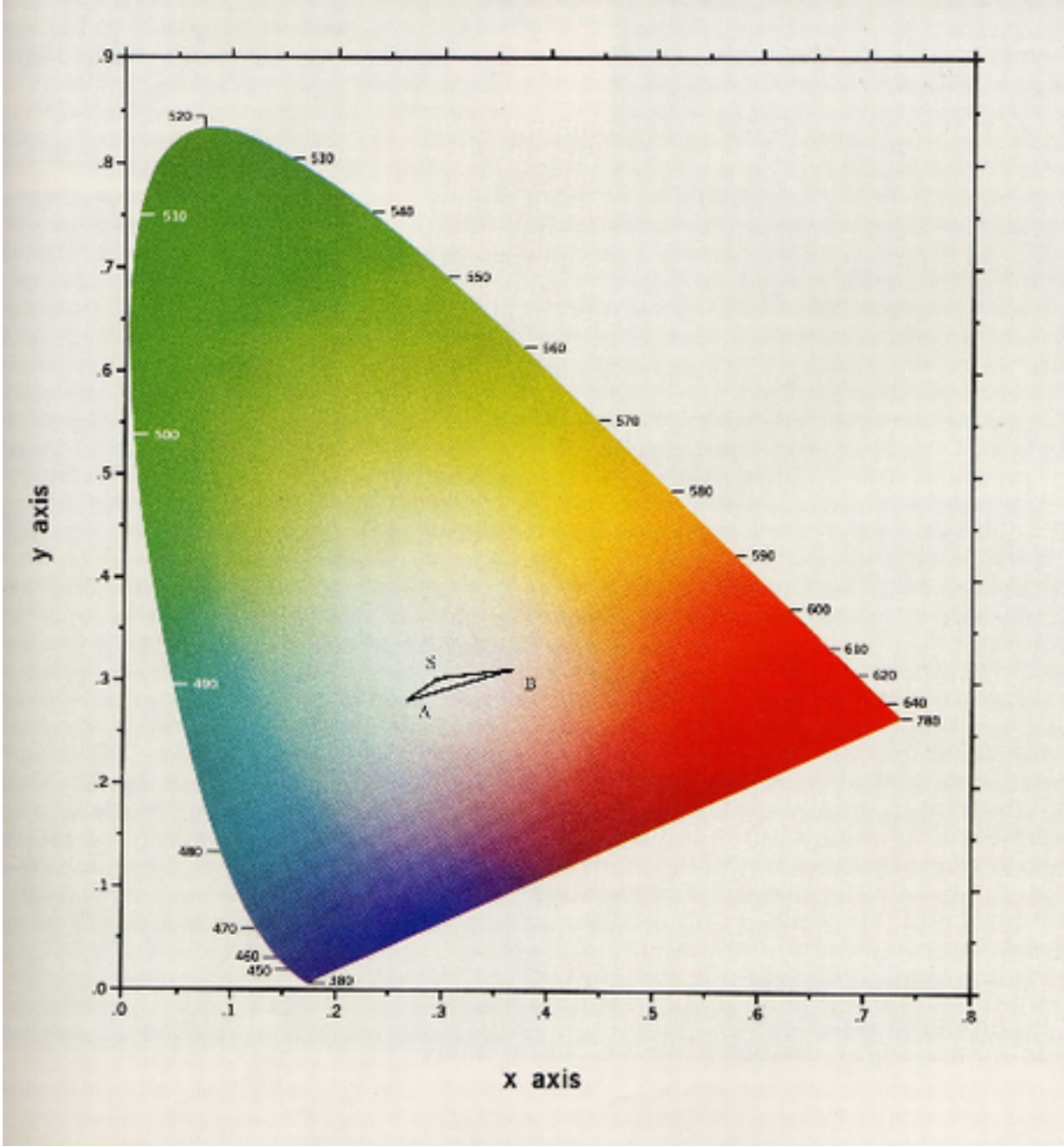


According to the figure above:
Hue of S: H(S) $\approx$489
Hue of A: H(A) $\approx$486

(d) Rank the light sources with respect to their saturation S.
With B is the reference white point:

I think that saturation of A > saturation of S because its closer to the edge (where the dominant wavelength is and so the maximum saturation.)
S(A) = dist{ A, B} / dist{ H(A), B} > S(S) = dist{ S, B} / dist{ H(S), B}
Is this the correct answer?
Yes

(e) Plot the region of colors which is produced through the mixture of S, A and B.
Triangle between S, A and B.



Exercise 2

We consider the representation of colors in a color space. In Figure A.1 (see attachment), the color matching functions of the CIE X, Y and Z primary colors are given. Further, in table 1 (see attachment) their spectral values are given with 10 nm interval (e.g. the spectral color of 500 nm has the following tri-stimulus values $\bar{x}$= 0.0049, $\bar{y}$= 0.323 and $\bar{z}$= 0.2720). Given a color K(λ) with a certain spectral distribution, then $X = \int_{[?][?]} K([?][?])\,\bar{x}([?][?])d[?][?]$, $Y = \int_{[?][?]} K([?][?])\,\bar{y}([?][?])d[?][?]$ and $Z = \int_{[?][?]} K([?][?])\,\bar{z}([?][?])d[?][?]$ It is assumed that K(λ) is a white light source i.e. equal energy distribution over all wavelengths.

(a) Compute X, Y and Z for a given color A of 500 nm. Further, calculate the chromaticity coordinates x = X/(X+Y+Z) , y = Y/(X+Y+Z) and z = Z/(X+Y+Z) of A.

$$K(\lambda) = k$$

$$X = k \int_{\lambda} \bar{x}(\lambda) d\lambda$$

$$X = k \int_{\lambda} \rho(\lambda) \cdot \bar{x}(\lambda) d\lambda \qquad \rho(\lambda) = \rho(\lambda_{500})$$

$$X = k \cdot \rho(\lambda_{500}) \cdot \bar{x}(\lambda_{500}) \qquad \rho(\lambda_{500}) = 1$$ (because $K(\lambda)$ is a white light source)

$$X = k \cdot \bar{x}(\lambda_{500})$$

$$X = k \cdot 0.0049$$

similar for Y and Z, so:

$$Y = k \cdot 0.323$$

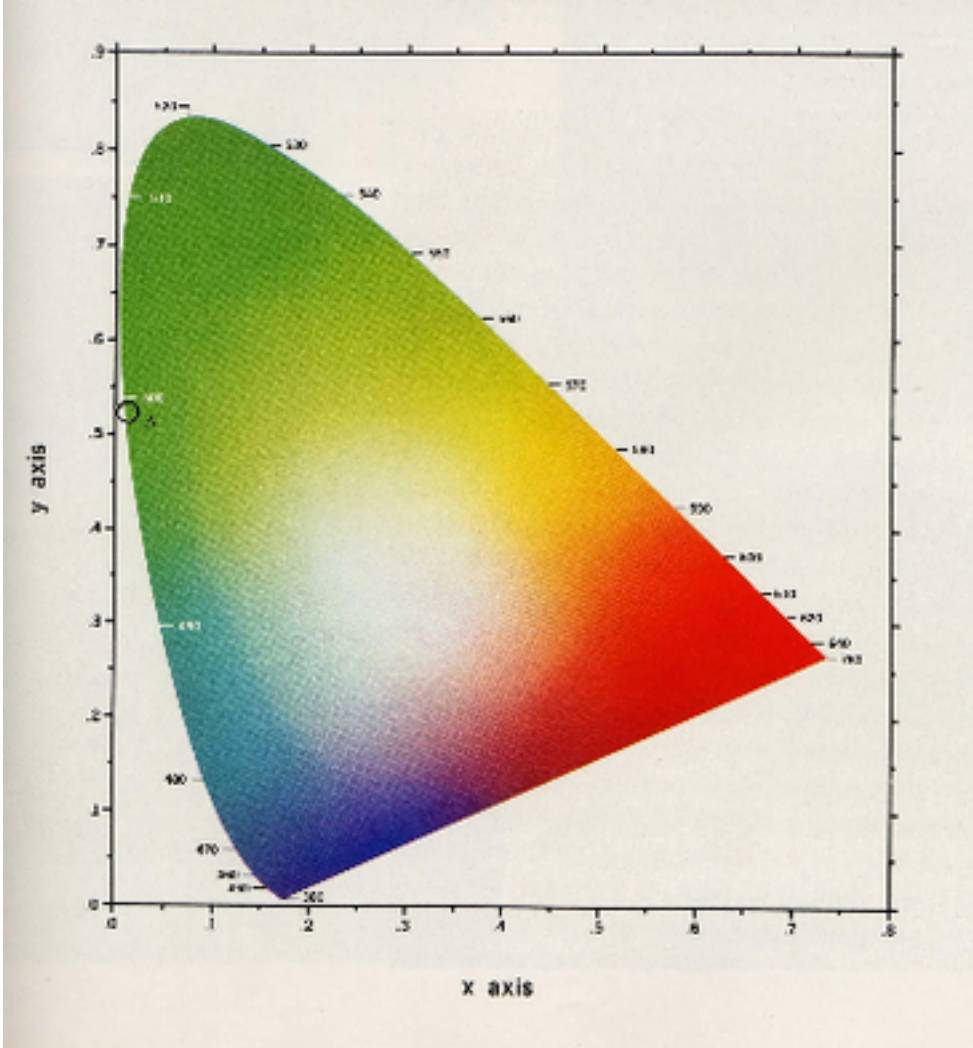
$$Z = k \cdot 0.272$$

$$x = \frac{k \cdot 0.0049}{k \cdot (0.0049 + 0.323 + 0.272)} = 0.0081$$

$$y = \frac{k \cdot 0.323}{k \cdot (0.0049 + 0.323 + 0.272)} = 0.538$$

$$z = \frac{k \cdot 0.272}{k \cdot (0.0049 + 0.323 + 0.272)} = 0.453$$

(b) Plot color A as a small circle in the chromaticity diagram given in Figure 2.



(c) Given a color B of 580 nm, find X, Y and Z and the chromaticity coordinates x, y en z.

$X = k \cdot 0.9163$

$Y = k \cdot 0.8700$

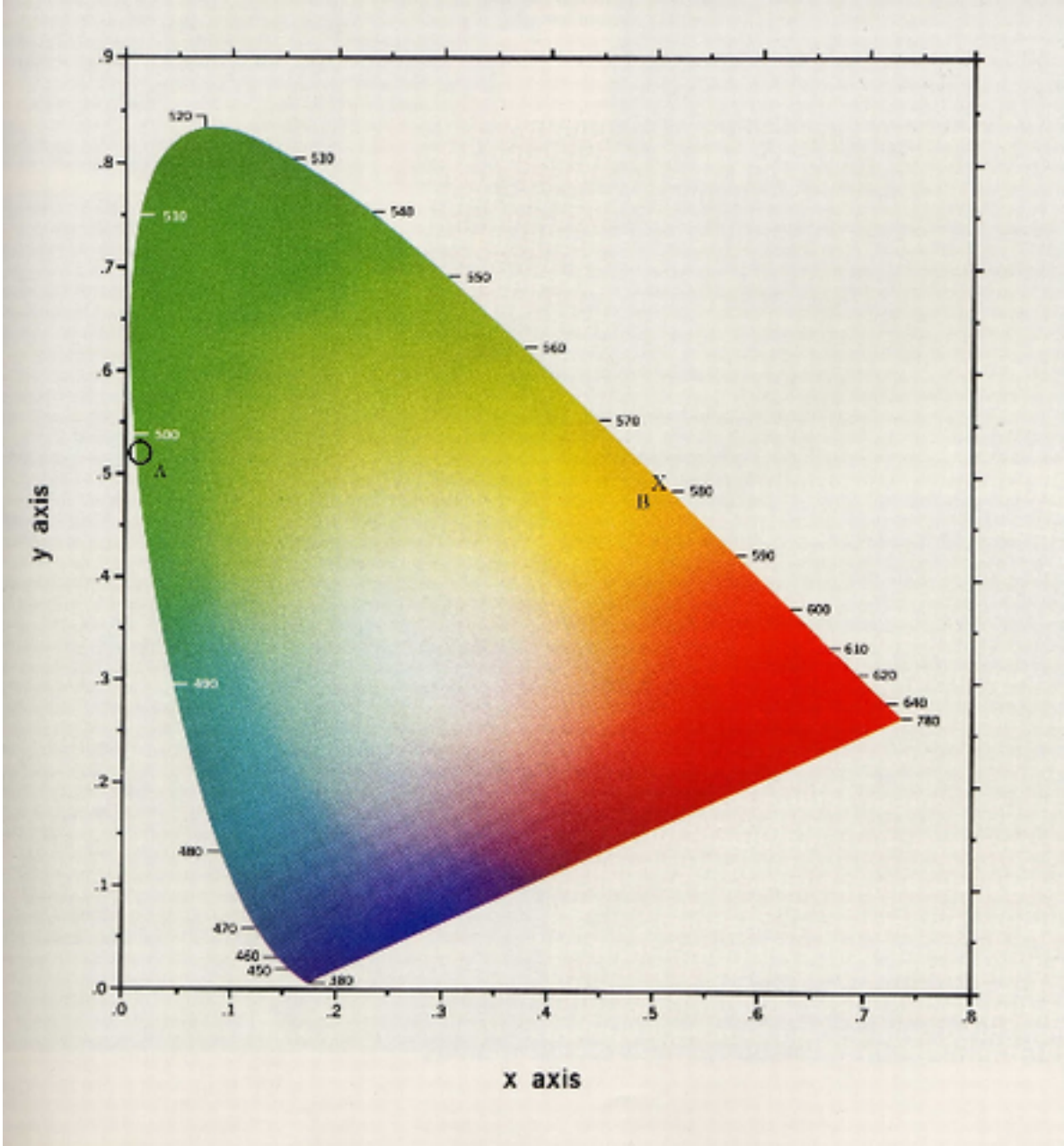
$Z = k \cdot 0.0017$

x=0.512

y=0.486

z=0.0

(d) Plot color B as a small cross in the chromaticity diagram.



(e) Given a color C consisting of the colors A of 500 nm and B of 580 nm. Compute X, Y and Z and the chromaticity coordinates x, y en z.

X = $k.(0.0049 + 0.9163)$

Y = $k.(0.323 + 0.87)$

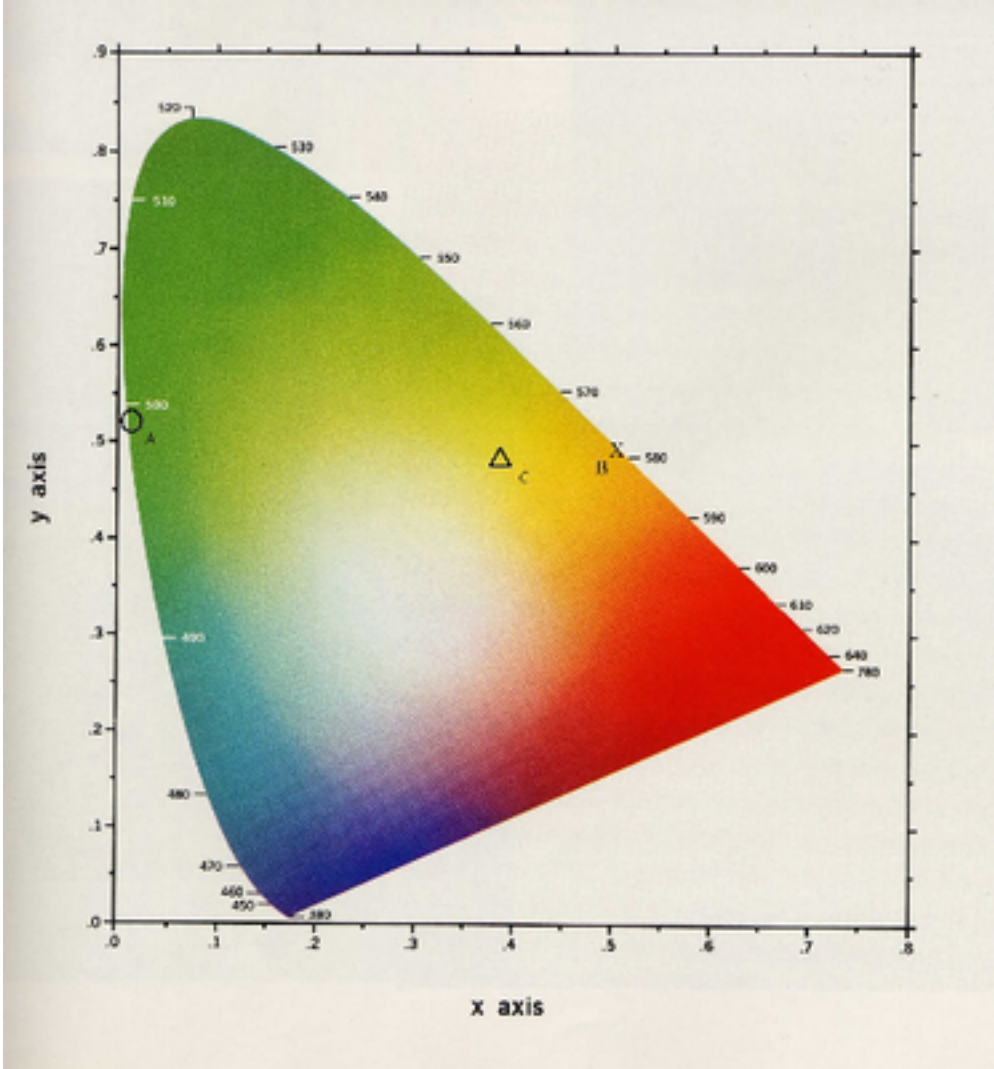
Z = $k.(0.272 + 0.0017)$

$$x_C = \frac{k\cdot(0.0049+0.9163)}{k\cdot((0.0049+0.9163)+(0.323+0.87)+(0.272+0.0017))} = 0.385$$

$$y_C = 0.499$$

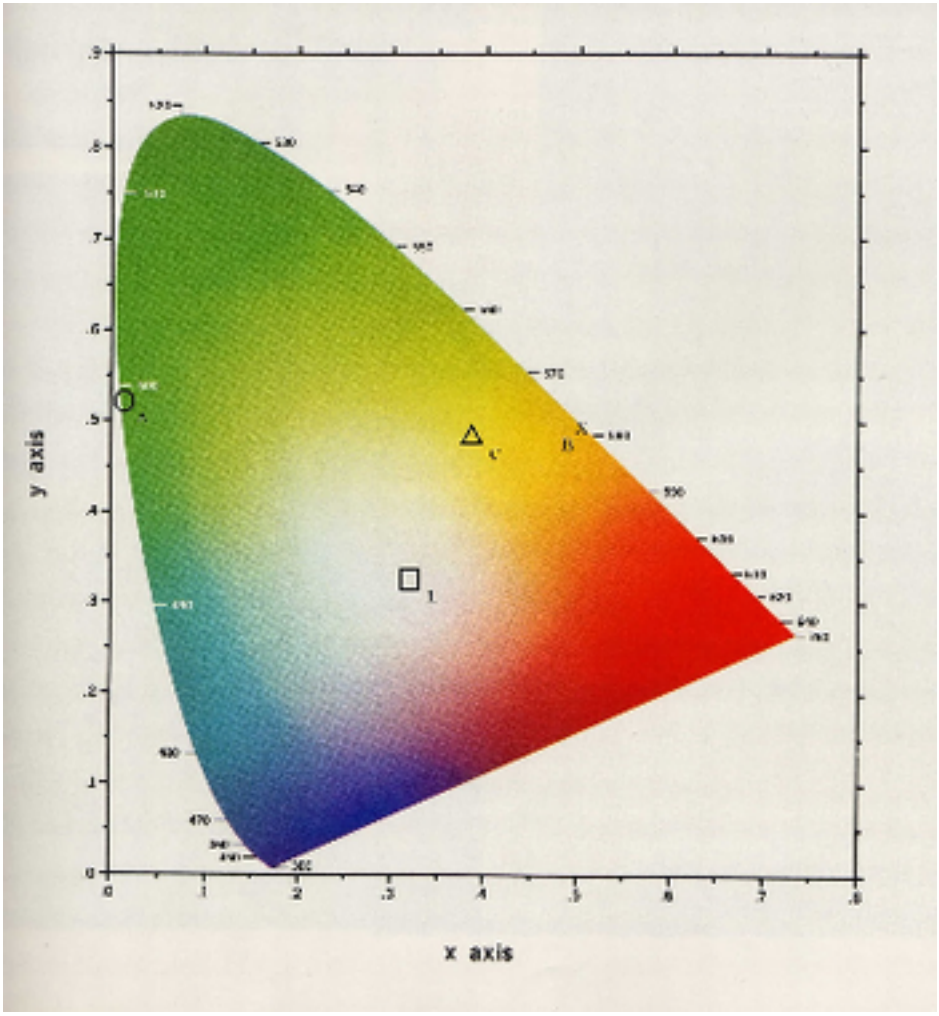
$z_C = 0.115$

(f) Plot the color as a small triangle in the chromaticity diagram.



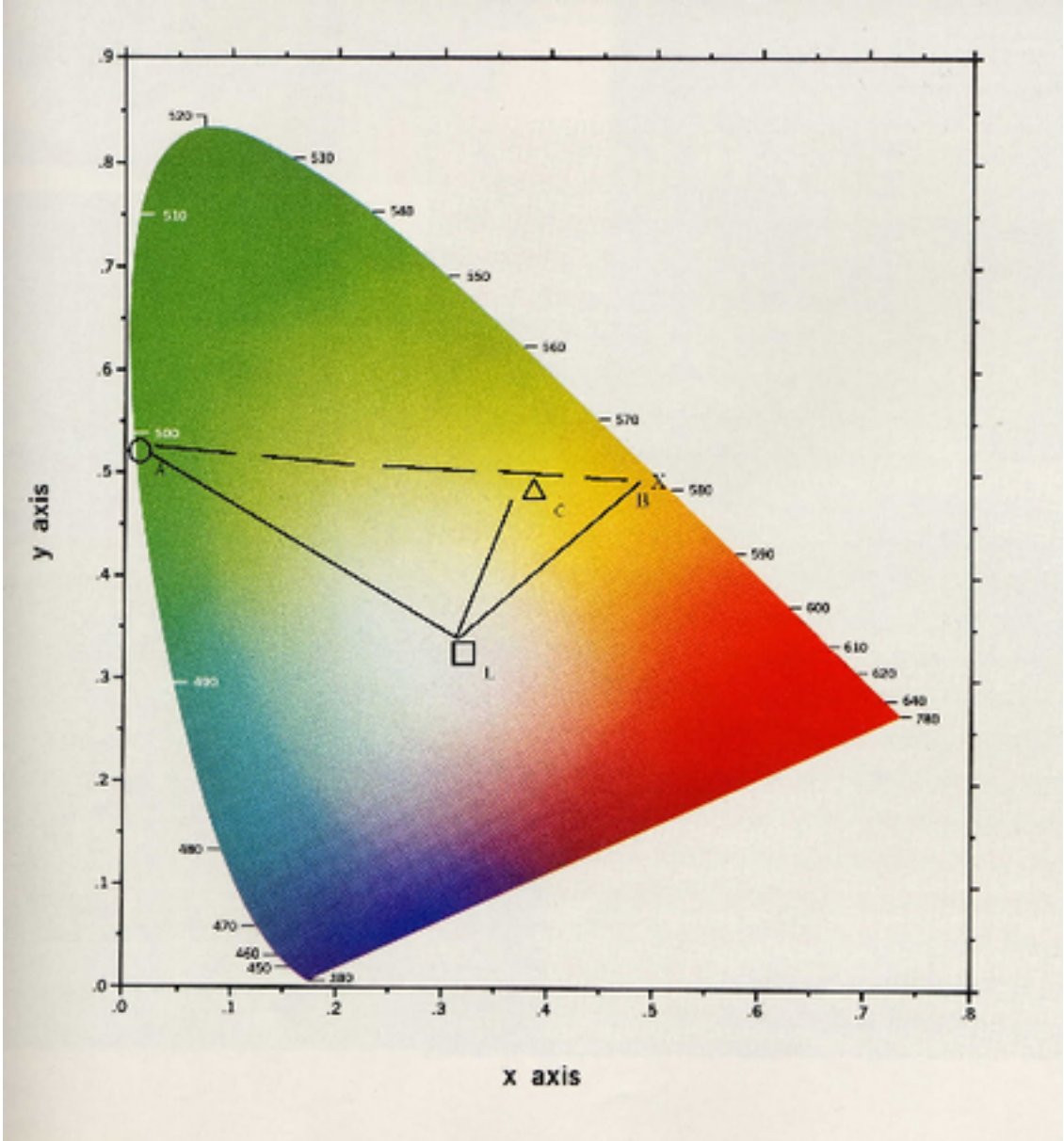
(g) If the white light source K(λ) varies (only) in intensity what would happen with the values X, Y , Z and x, y and z of the colors A, B and C? What will be the consequence?
X,Y and Z change with K
x, y and z invariant
A,B and C remain the same

(h) The tri-stimulus values of a given lamp L are as follows X = 98.04, Y = 100.00 and Z = 118.12. Compute the chromaticity coordinates x, y and z and plot color L with a small rectangle in the chromaticity diagram.
x=0.31
y=0.316
z=0.373

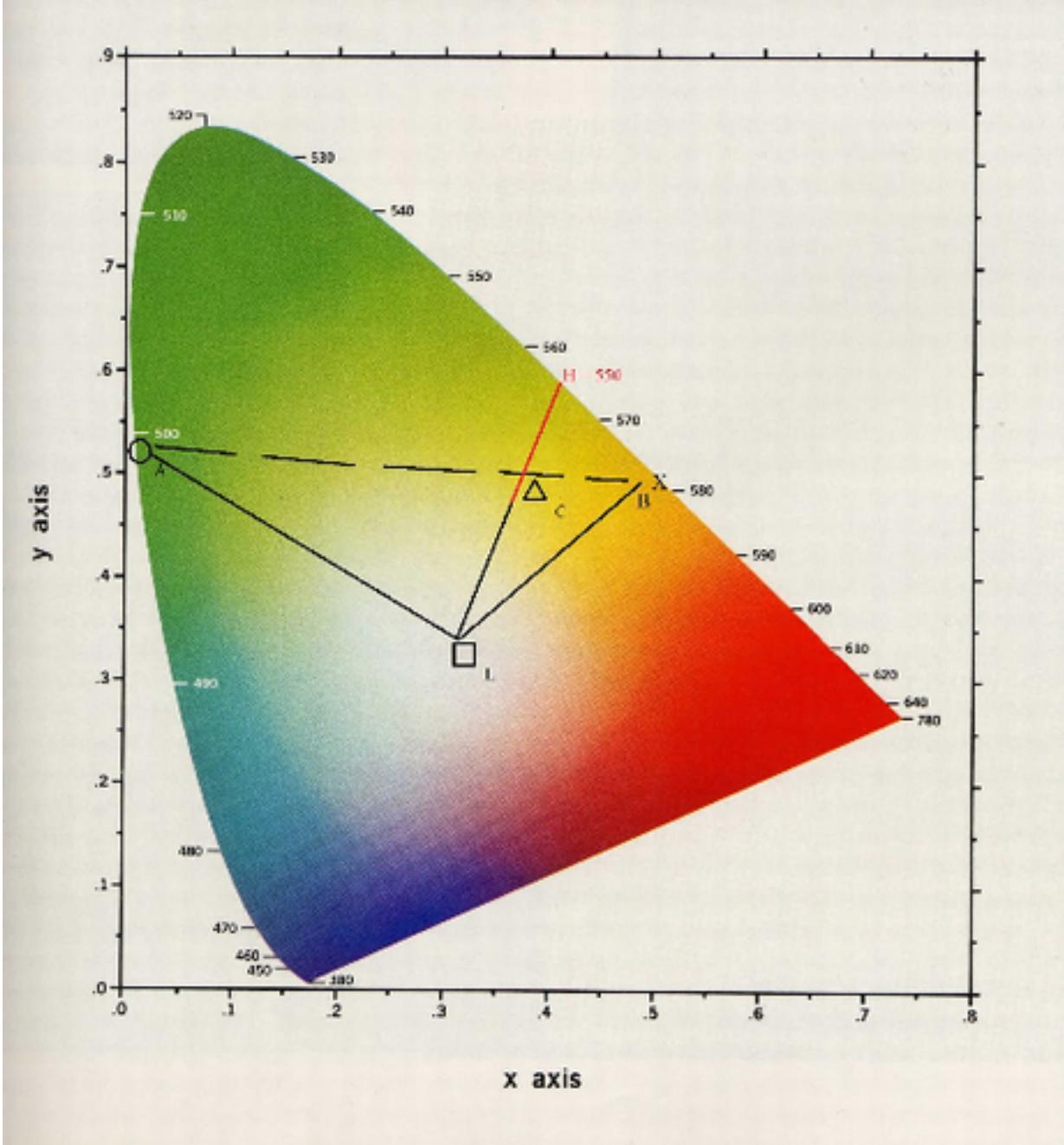


(i) Indicate, by three different lines, the colors which are generated by the mixture of L with A, B and C respectively.

Lines between L and A,B&C



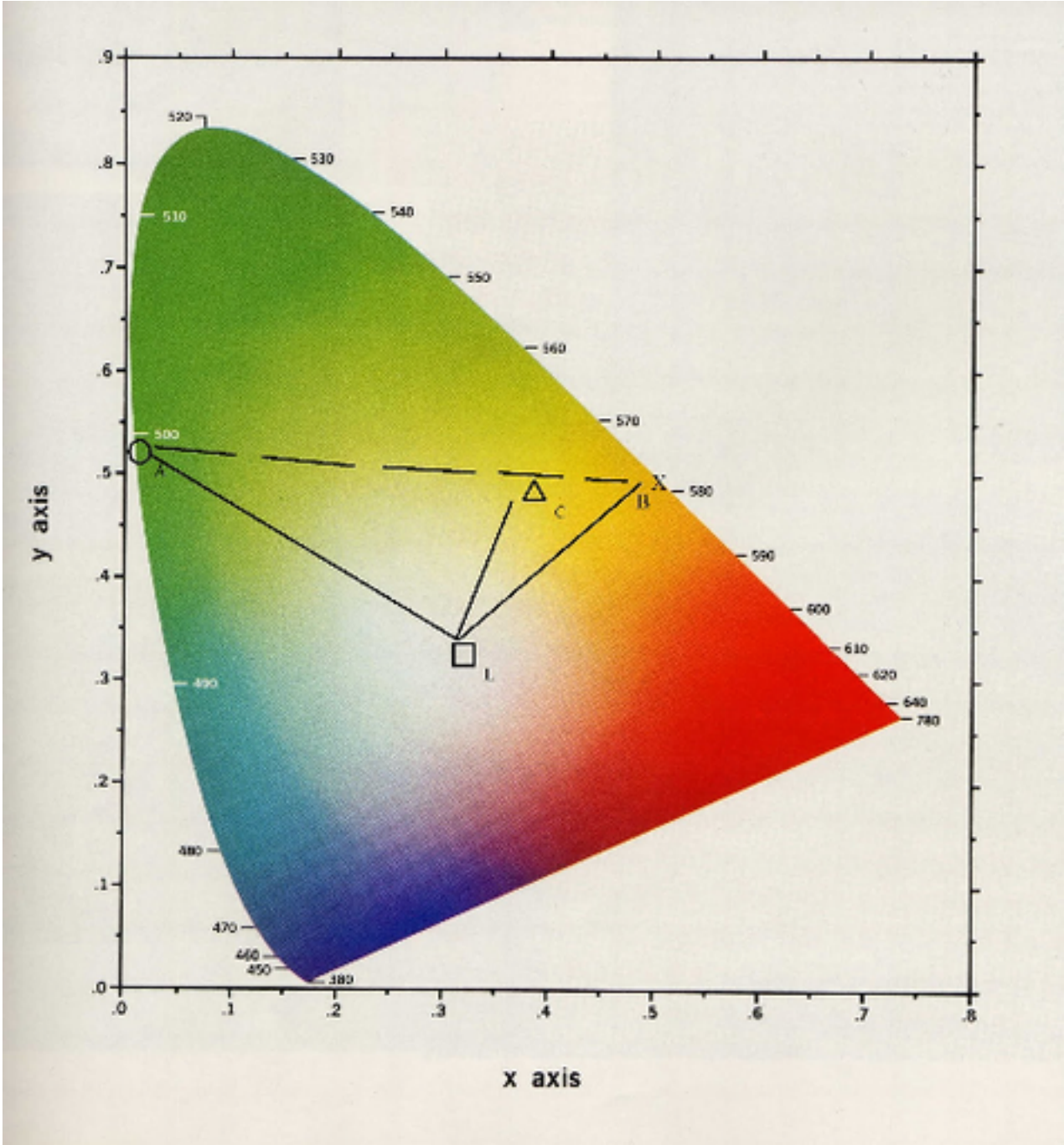
(j) What is the hue (dominant wavelength) of C with L as reference white?
Hue of C 540nm. How? According to the figure below, its 564nm.
But average of 500 and 580 (the two constituent colours) is
540.



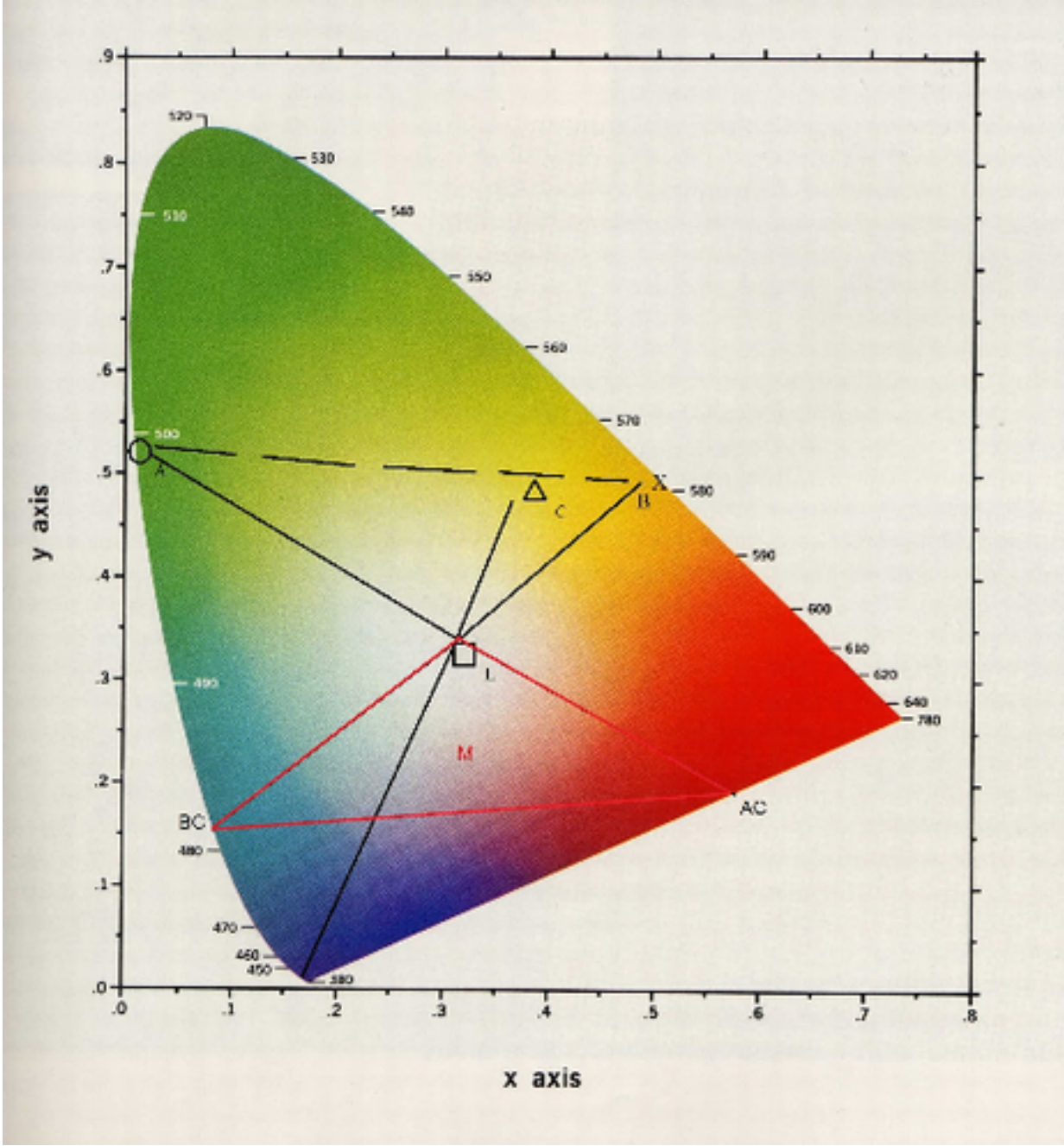
(k) Order the three colors A, B and C with respect to their saturation.
(most) A=B, C (least)

(l) What are the complementary colors A^c, B^c, and C^c for A, B and C respectively with L as reference white? Are these complementary colors pure (wavelength) or a mixture of pure colors?
C^c and A^c are mixtures of pure colours (purple) B^c is a pure colour.

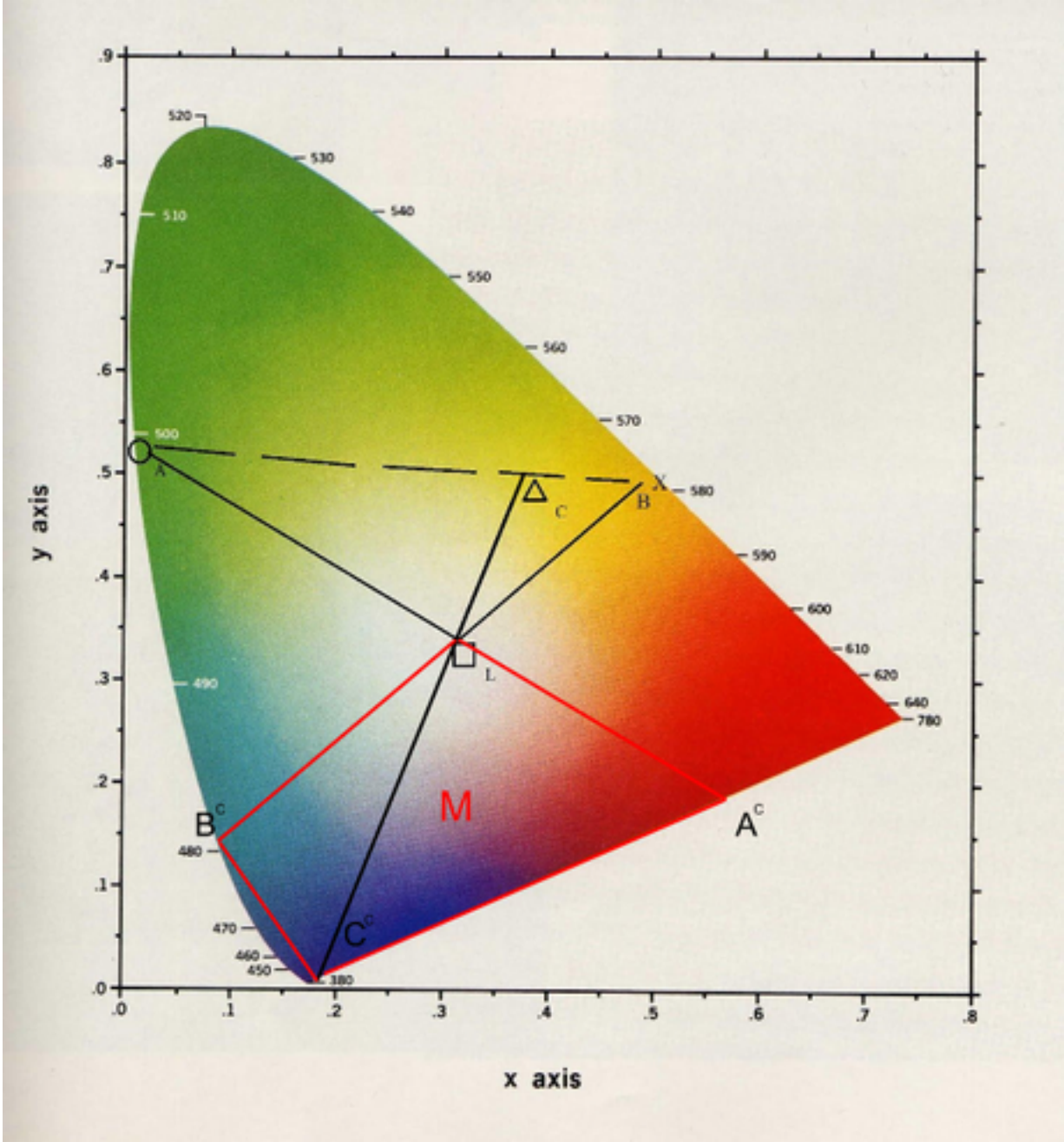
-> mirror dominant wavelength through reference light, if the complementary wavelength is on the spectral line, it is a pure color, if its on the “bottom” line its a mixture of colors.



(m) Draw the region of colors which are generated by the mixture of A^c, B^c, C^c and L.



Shouldn’t it be this:



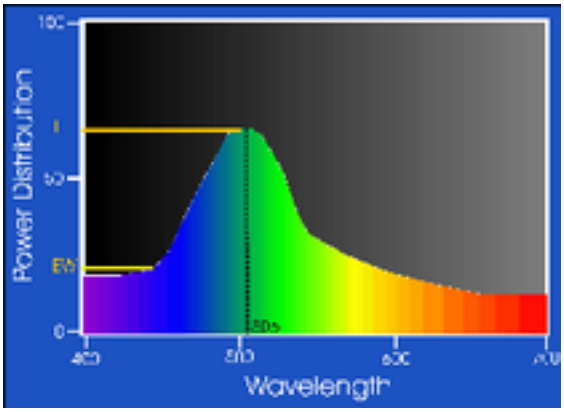
I think that the region must be a triangle but i may be wrong, yes you are [wrong]!
All colors that can be created using three light sources, lie in a triangle. The mixtures of four light sources, however, lie in a quadriliteral.

(n) Given is a color with a spectral power distribution given in Figure A.2 (see attachment). Estimate the hue (dominant wavelength) and describe the amount of the saturation and intensity. What should be the approximated position of this color in the chromaticity diagram?

$\approx 510\ nm$

Does anyone know how to describe the amount of saturation and intensity?

I'm not sure on the formal notation, however considering the graphs i think it should be done like slide 64 of lecture 2:



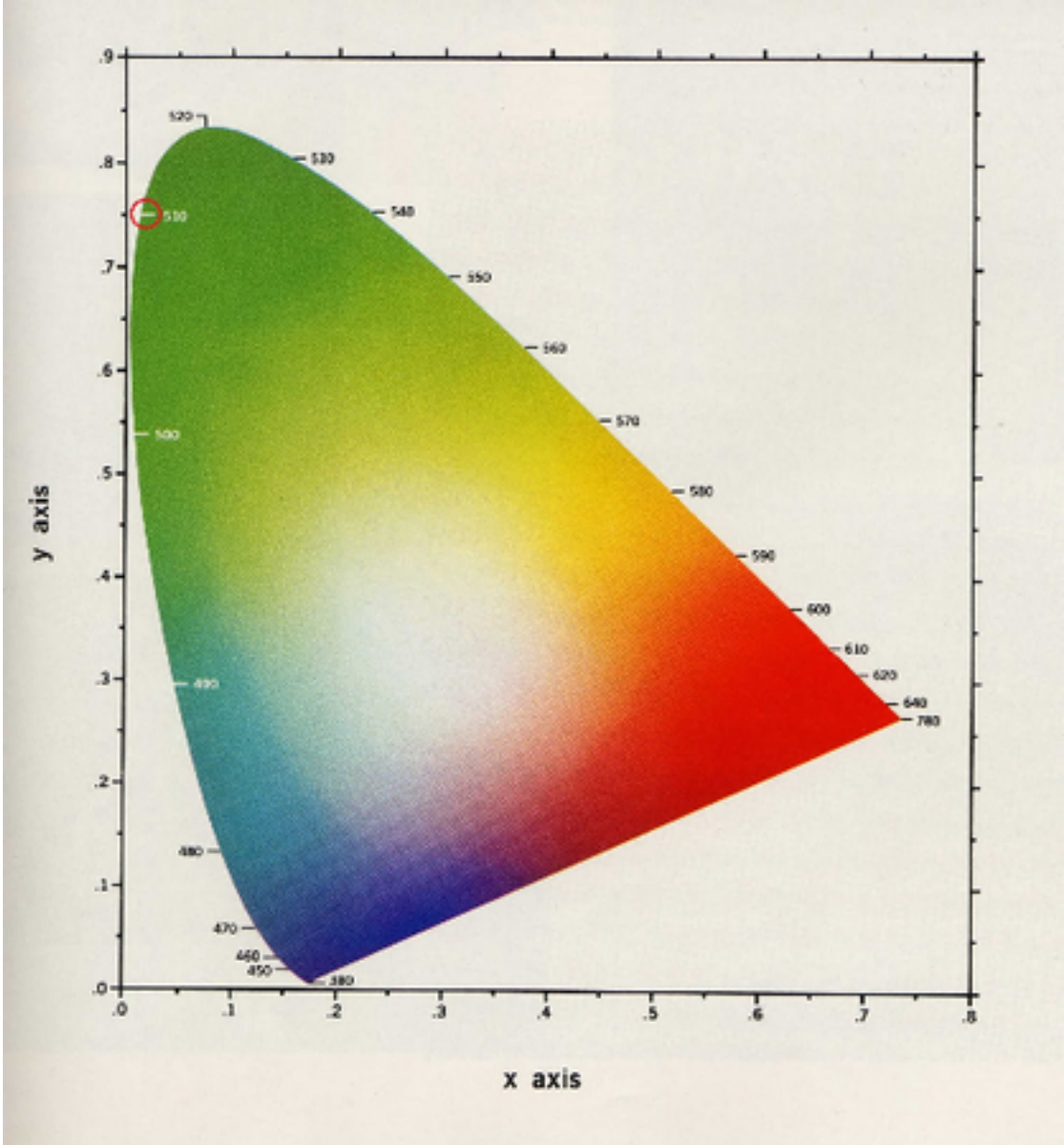
The top yellow line represents EH, the bottom EW;

Hue: dominant wavelength of the SPD: EH

Saturation: purity of the colour: EH-EW

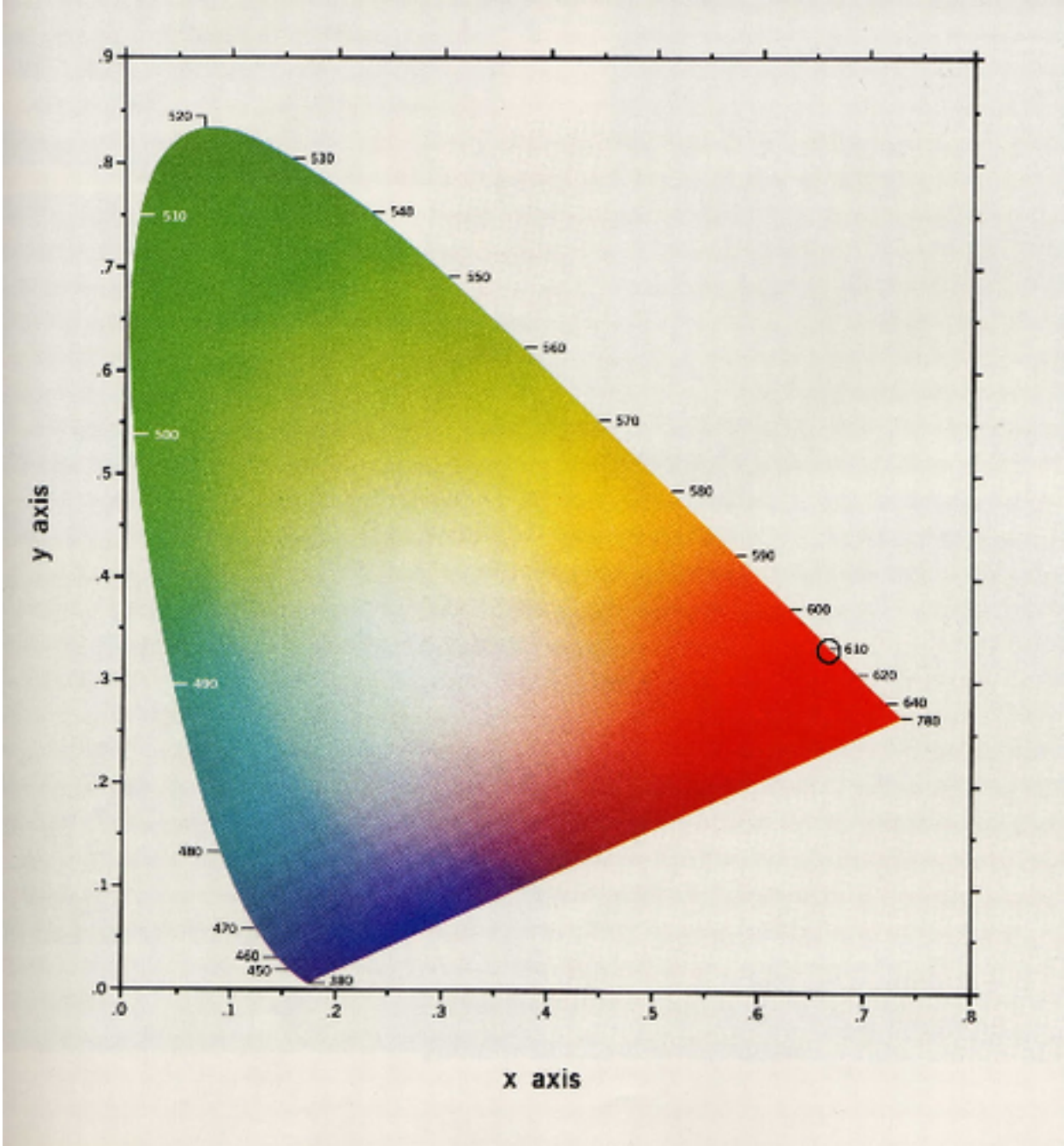
Intensity: brightness of the colour: EW

The graphs however are not labeled on the sheet, so perhaps a estimation/review of the graphs..



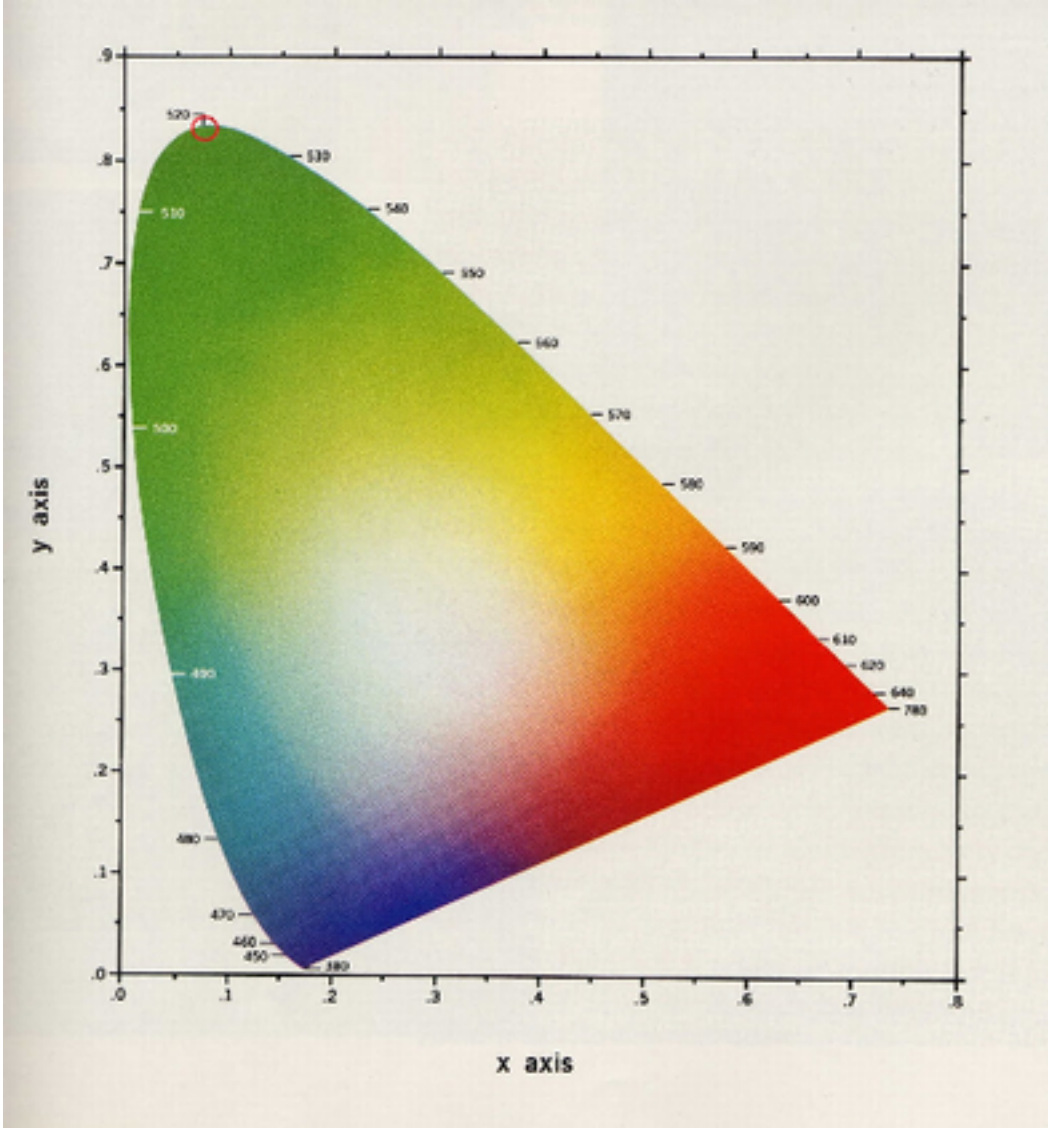
(o) Given is a color with spectral power distribution given in Figure A.3. Estimate the hue (dominant wavelength) and describe the amount of the saturation and intensity. What should be the approximated position of this color in the chromaticity diagram?

$\approx 610\,nm$



(p) Given is a color with spectral power distribution given in Figure A.4. Estimate the hue (dominant wavelength) and describe the amount of the saturation and intensity. What should be the approximated position of this color in the chromaticity diagram?

$\approx 520\,nm$



(q) For which of the three spectra a human will perceive the highest intensity? Explain your answer.

See figure A1. Answer A3 (2o).

I am stupid. Please answer more explicitly.

He said that its the green spectra which is the y in figure A.1

The human eye perceives green as being more intense due to the number of green sensitive cones.

I believe it is A3, the colour green because the dominant wavelength (~600) has a high rating in figure A1, as where the other wavelengths do not have a high value in the plot.

Exercise 3

We consider the color of a matte, dull (not glossy) surface. The color at a specific location on the surface under white light illumination is given by the following simple reflection model

$R = I\,k_R\,cos(\theta)$, $G = I\,k_G\,cos(\theta)$ and $B = I\,k_B\,cos(\theta)$, where I is the intensity of the white light source,
k_R, k_G and k_B are the amount of red, green and blue reflected by the surface (i.e. color of the surface). Furthermore, $cos(\theta) = \vec{n} \cdot \vec{l}$ is the dot product of the two-unit vectors $\vec{n}$ (i.e. surface normal) and $\vec{l}$ (i.e. direction of the light source), see Figure A.5.

(a) Assume that the surface is flat and homogeneously colored. Explain why the intensity is higher when the surface normal coincides with the direction of the light source than observed under an angle with respect to the direction of the light source.

$R = I \cdot k_R \cdot cos(\theta)$
$G = I \cdot k_G \cdot cos(\theta)$
$B = I \cdot k_B \cdot cos(\theta)$
$cos(\theta) = \pi \cdot L$
$I = constant$
k_R, k_G, k_B are constant
variations due to θ

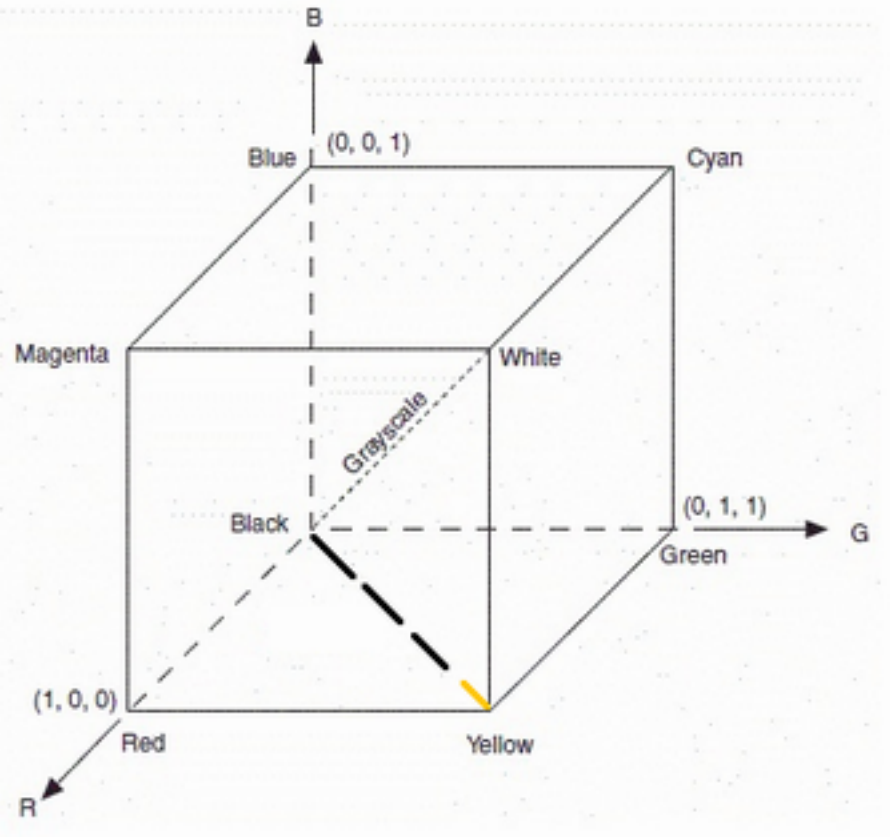
When the surface normal coincides with the direction of the light source theta = 0, so cos(theta) = 1. R, G and B are highest when theta=0.

Intensity = (R+G+B)/3
Intensity is highest when R,G and B are highest

(b) Assume that the color of the surface is yellow i.e. R = 100, G = 100, and B = 10. Explain what will happen with the values R, G and B if (only) the intensity of the light source will diminish. Plot the positions of the colors in the RGB-color space.
They will remain relatively the same, variations are in I

Normal	Diminished
$R = I \cdot k_R \cdot cos(\theta)$	$R = \frac{1}{n} I \cdot k_R \cdot cos(\theta)$
$G = I \cdot k_G \cdot cos(\theta)$	$G = \frac{1}{n} I \cdot k_G \cdot cos(\theta)$
$B = I \cdot k_B \cdot cos(\theta)$	$B = \frac{1}{n} I \cdot k_B \cdot cos(\theta)$

Since RG and B depend on the intensity, I think that they will decrease too.
Ratio between RGB will stay the same, values will decrease.
How do I plot this in the RGB color space? It gets very unclear very easily in the 3D cube...
I think that you have to draw the point 100,100,10 in the rgb-cube and then draw a line from this point to 0,0,0 where intensity will be zero.
^^ Are you a Eugenio?



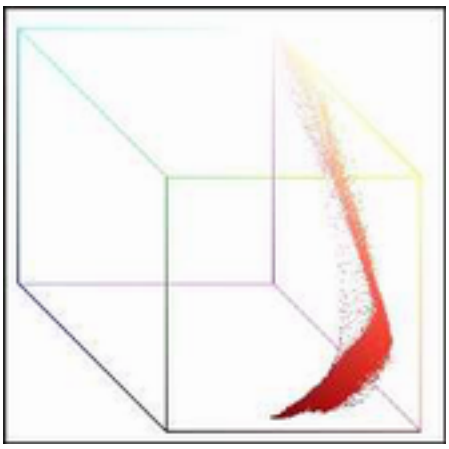
(c) In case of a curved (not flat) surface, indicate where the colors will be positioned in the RGB-color space. Explain your answer.

Colour doesn't change, only intensity changed

- Intensity changes I
- Shading (curved surface)
- Shadows

From lecture 3 slides page 15 is an example of RGB color space representation of a color on a not matte surface.

This is for glossy surfaces, we have a matte here. I think that, the line will be the same as previous. In a curved surface, theta changes, intensity is maximum in theta 0 and minimum in theta 90. We perceive different intensity from Imax to Imin from different points in the curved surface.



(d) A simple color invariant is given by R/G. Proof that R/G is independent of the (intensity) light source I, object geometry and the direction of the light source.

$$R = I \cdot k_R \cdot cos(\theta)$$

$$G = I \cdot k_G \cdot cos(\theta)$$

$$\frac{R}{G} = \frac{I \cdot k_R \cdot cos(\theta)}{I \cdot k_G \cdot cos(\theta)} = \frac{k_R}{k_G}$$

Independent of intensity and angle because they cancel out.

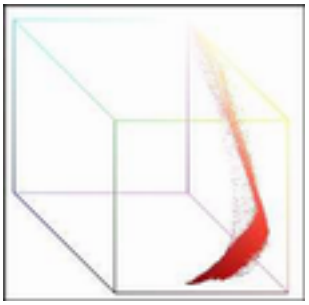
(e) The values of R, G and B will vary for a curved surface. Give the approximated shapes of the histograms for a homogeneously (curved) surface for R, G, B and R/G. Which color models will you choose for the recognition of objects under varying light intensity. Explain your answer.

R,G,B curved

$\frac{R}{G}$ single peak

Normalized rgb because it is invariant to intensity

(f) Consider the same surface. Assume that the surface is glossy (instead of matte). The reflection model is now given by $R = I\,k_R\,cos(\theta) + I\,k_S\,cos^{\,n}(\alpha)$, $G = I\,k_G\,cos(\theta) + I\,k_S\,cos^{\,n}(\alpha)$ **and** $B = I\,k_B\,cos(\theta) + I\,k_S\,cos^{\,n}(\alpha)$. k_S **is the specular reflection coefficient and cosn depends on the glossiness and α depends on the viewing condition. Plot the colors of the homogeneously colored (shiny) surface in RGB- and rgb-color space.**



...

What the fuck is the rgb color space? It is not mentioned in any of the papers I'm pretty sure.

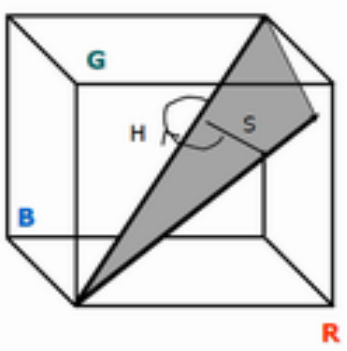
Successful Troll is successful

think with rgb color space he means this chromaticity diagram ..

-->if you try to calculate r = R/(R+G+B) and g = G/(R+G+B) the intensities cancel out.

So the equations depend on $cos^n(\alpha)$ and $cos(\theta)$ **they (theoretically) range from -1 to 1, so i assume for the chromaticity diagram it will result in an area, depending on constants** k_G k_B k_r **and** k_s **?**

Following this logic, in the RGB cube it would be a volume..



$$H(R,G,B) = \arctan\left(\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right)$$

We have the same line as before for matte surfaces. We can draw the same line as before for the matte surface and just draw a hyperplane on this.

However, if anyone knows a proper answer to this just share it...

(g) Proof that $\frac{R}{G}$ **is not a color invariant for shiny surfaces. Proof that** $\frac{R-G}{R-B}$ **is a color invariant for shiny surfaces.**

$$\frac{R}{G} = \frac{I \cdot k_R \cdot cos(\theta) + I \cdot k_S \cdot cos^n(\alpha)}{I \cdot k_G \cdot cos(\theta) + I \cdot k_S \cdot cos^n(\alpha)}$$

not invariant.

Invariant!

Exercises II

Exercise 1

a.
A uniform filter replaces each pixel with an average of its neighbourhood. If we increase the size of the neighborhood we will lose details, making the image blurrier. So, with a 7x7 uniform filter we will obtain a blurrier image.

b.
The following 3x3 edge filter will highlight Vertical Edges:
-1 0 1
-1 0 1
-1 0 1

c.
Uniform filter =
1 1 1
1 1 1
1 1 1

Laplacian filter =

1 -2 1
-2 4 -2
1 -2 1

$f \otimes g$ (convolve the laplacian filter with the uniform filter)

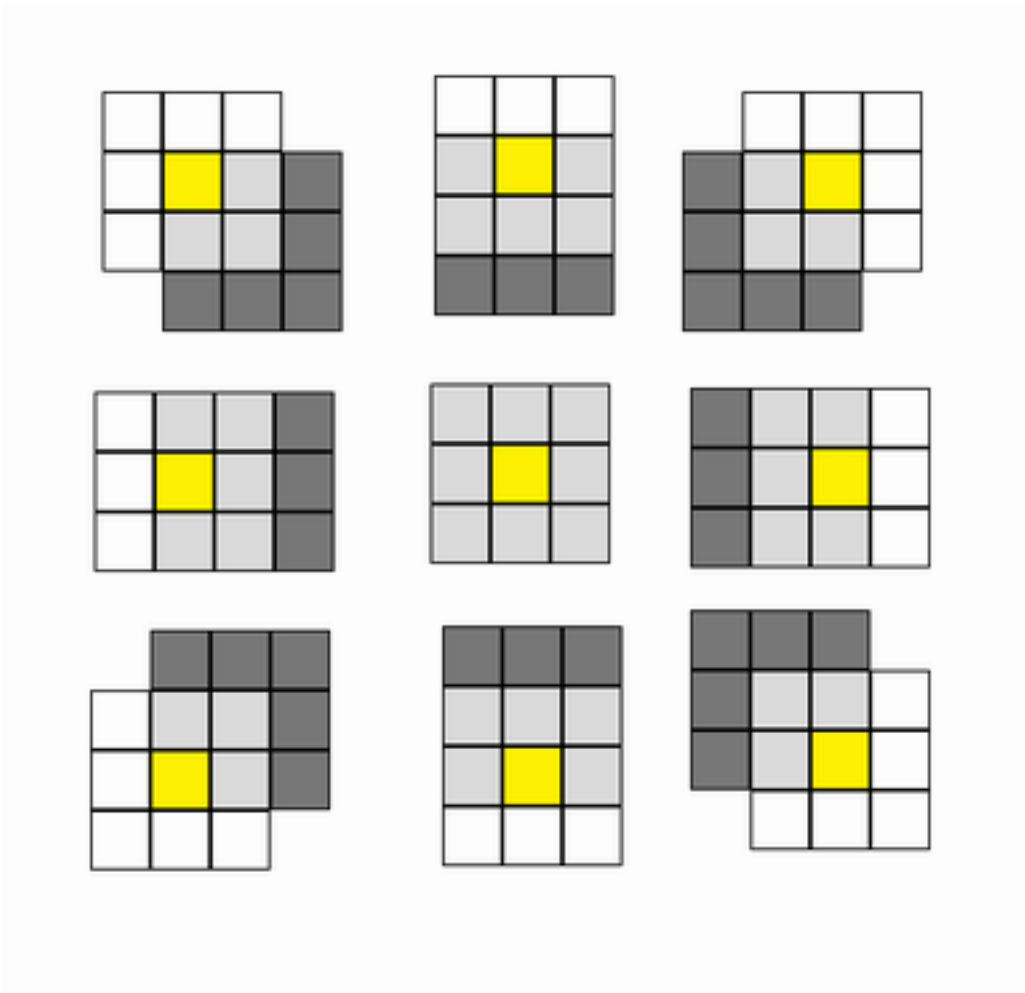
1	-1	0	-1	1

-1	1	0	1	-1
0	0	0	0	0
-1	1	0	1	-1
1	-1	0	-1	1

Does anyone know a good link to a step by step example ?

Just shift everything! :D

No need to fear, convolution is here!



Exercise 2

a.

0	20	0	0
0	20	0	0
0	20	0	0

To compute the last column, make a copy of the last column (there are other ways of doing this)

b.

85	85	85	85
85	85	85	85
85	85	85	85

⅓ * 255 = 85

Why do we multiply it with 255? The normalized RGB is between 0 and 1. Right?
But teacher told to do so. Its called scaling it back up.

c.

0	0	0	0
0	0	0	0
0	0	0	0

d. How can you classify the transition to be of a shadow type?
Edge response present in RGB but not in normalised RGB.

e. With the same procedure, could you distinguish shadow edges from geometry edges? Please explain.
No, shadow steeper, rough. Shading is smooth, use the profile of Edges.

f. With the same procedure, how can highlights be classified? Do you need more than two color features? Which ones?
Possible, compare c1c2c3 and l1l2l3.
We need the hue to distinguish highlighted edges.
The color features needed are: hue and edge maps.

g. Edge classification can be used as a pre-process for image/video retrieval. What are the remaining steps?
Find interesting points, extract descriptors (at salient points), Edge and corners
Detect the descriptors such as salient points, and interesting points such as edges and corners.

h. For image and video retrieval, do you prefer to use local descriptors or global descriptors? Explain why.
Global: very sensitive to occlusion, change of viewpoint, computationally cheap.
Local: can still find objects even if occluded.

Exercise 3

a.Give an example of an image for which the white patch method will fail. Please explain.
Examples: jungle pictures, sunset or any other biased pictures.
b. Explain in words how these color constancy methods work.
White patch: find the highest value in the image (max R,G & B), assume it's white and adjust the image.
Grey World: average of everything is grey.

c. White patch :

$$a = \frac{255}{max\{R\}} = \frac{255}{180}$$

$$b = \frac{255}{230}$$

$c = \frac{255}{220}$(per channel), $c = \frac{255}{120}$ (per pixel) why per channel and per pixel?

Grey world :

$a = \frac{128}{\overline{R}}\,\overline{R}$ is the average R from the patch (sum of all pixel values divided by the total number)
We find the average values of the image's R, G, and B color components and use their average to determine an overall gray value for the image, gray_value, e.g. 128 (in the example it'08.85). Each colors actually 1 component is then scaled according to the amount of its deviation from this gray value. We obtain the scale factors by dividing the gray value by the average of each color component. I.e. c = scale_B = gray_value / avg(B).

avg(R) = 140.000

→ scale all red pixel values by (108.889 / 140.000) = 0.778
avg(G) = 91.111
→ scale all green pixel values by (108.889 / 91.111) = 1.195
avg(B) = 95.889
→ scale all blue pixel values by (108.889 / 95.889) = 1.136
AREN'T WE SUPPOSED TO ASSUME AVERAGE PIXEL VALUE TO BE 128 ALWAYS, EVEN IF THE REAL AVERAGE IS SOMETHING ELSE?

d. What is the grey-edge assumption?
The grey-edge assumption, assumes that the average edge difference in a scene is achromatic (grey).

e. **Give an example of an image when the grey-edge assumption does not hold.**
An image that has a few colored patches.

f) What are natural image statistics and how can they be used for color constancy?
If there are many edges, use them etc. According to the statistics we can choose a suitable method.
We fit a symmetric Weibull distribution on the edge distribution. The Weibull distribution is defined by its two parameters gamma and beta which denote the amount and size of edges in the image. These parameters will have values in a certain range depending on if it's e.g. a nature or city image. So then we can use a suitable color constancy method depending on the kind of image. For example a forest image will likely not contain a grey edge.
So what are natural image statistics? Definition please?
With natural image statistics we can extract the most important characteristics of color images

Exercise 4

a.
Take the square root of the sum of the squares of [per component: the partial derivative for the component times the uncertainty of that component].

$$I = R + G + B$$

Say $\sigma R = \sigma G = \sigma B = 4$. dI / dR is 1, etc.

$$\sigma I = \sqrt{(1*\sigma R)^2 + (1*\sigma G)^2 + (1*\sigma B)^2} = \sqrt{48}$$

Independent of RGB, only dependent on noise.

b. $$\frac{1}{R} = \left(\frac{0*R - 1*1}{R^2}\right)*\sigma R = \frac{-\sigma R}{R^2}$$

It is not stable and depends on the noise and intensity of the channel.

c. Under which circumstances do you think that normalized color and hue will become unstable?

The hue becomes unstable with low saturation of the color, because the hue is angular around the linear saturation axis (see HSL/HSV models). A certain hue difference in degrees will cause a larger jump in perceivable color when the saturation is lower.
Normalised colour is unstable with low intensity (see slides lecture 3)

$$\delta_r = \frac{\sqrt{R^2(\delta_B^2 + \delta_G^2) + (B+G)\delta_R^2}}{(B+G+R)^4}$$

also because the sum of r, g and b should be 1. For example, 200/601 is similar to 201/601, but 1/4 is half the value of 2/4. While 201-200 = 2-1 = 1.

d. How can error propagation be used for histogram construction for image retrieval? Please explain.
With the help of Variable Kernel Density Estimation. E.g. compute uncertainty for normalized color components r and g. As kernel we use an Additive Gaussian noise kernel (per normalized color component). Then we let the scale (smoothing) parameter (bandwidth) of the kernel (per normalized color component) be the uncertainty of that normalized color component.

Exercise 5

a.

Recall S1 = $\frac{|R_a|}{|R|} = \frac{4}{10}$

Precision S1 = $\frac{|R_a|}{|A|} = \frac{4}{15}$

Recall S2 = $\frac{|R_a|}{|R|} = \frac{4}{10}$

Precision S2 = $\frac{|R_a|}{|R|} = \frac{4}{15}$

b. (precision = number of documents / position of the found document)
S1

$R_1 = 1$

Recall = $\frac{1}{10} = 10\,\%$

Precision = $\frac{1}{1} = 100\,\%$

$R_1 = 2$

Recall = $\frac{2}{10} = 20\%$

Precision $= \frac{2}{3} = 67\%$

$R_1 = 3$

Recall = $\frac{3}{10} = 30\%$

Precision = $\frac{3}{8} = 37.5\%$

S2

$R_2 = 1$

Recall = $\frac{1}{10} = 10\%$

Precision = $\frac{1}{2} = 50\%$

$R_2 = 2$

Recall = $\frac{2}{10} = 20\%$

Precision = $\frac{2}{6} = 33\%$

$R_2 = 3$

Recall = $\frac{3}{10} = 30\%$

Precision = $\frac{3}{11} = 27\%$

c.

R = 10 (number of relevant documents divided by number of correct results)

S1 = $\frac{3}{10}$

S2 = $\frac{2}{10}$

d.

depict differences for precision at rank 1,2,..,n for system A and B, PR_A(i)-PR_B(i) ?

^ A histogram for R-Precision itself is done over queries. So probably indeed he means a histogram for the difference of the Systems per Interpolated Precision (precision at rank 1,2,etc.)
TODO

Exercises III

Exercise 1

a. What is template matching and how can this technique be used for tracking?
Tracking is sliding window approach
Template is fixed and given beforehand

b. Could you define a pixel-wise similarity measure for template matching?
SSD-> sum of squared differences:
D(y) = sum over x in Omega ([I(x+y) - T(x)]^2 | minimize over y (error)
where: omega search window, T(x) value of pixel in target and I(x+y) value of pixel in target candidate (y is window offset).

Crosscorrelation:
C(y) = sum over x in Omega (I(x+y) * T(x) | maximize over y

c) What are the possible image transformations between the template and possible candidates? What is the search area?
Translation, rotation, scale, affine (like stretching)

d) What are the pros and cons of template matching for object tracking?
Pro: simple to implement, robust and accurate;
Con: time consuming ; only suitable for translation;

e) What is the difference between the similarity measure of template matching and mean-shift?
Mean shift uses histogram information instead of pixel values (its more flexible), faster in comparison to SSD.

Exercise 2

a) A vertical edge
b)
A =

1 1 1 0
1 1 1 0
1 1 1 0
1 1 1 0
$\vee f = \sqrt{f^2{}_x + f^2{}_y}$
f_x =
0 0 1 0
0 0 1 0
0 0 1 0
0 0 1 0
f_y=
0 0 0 0
0 0 0 0
0 0 0 0 if this is made with a vertical derative filter, why is there no bar of 1's?
0 0 0 0
$\vee f$ =
0 0 1 0
0 0 1 0
0 0 1 0
0 0 1 0

M (second moment matrix)=
$[f^2{}_x \; f_x f_y]$
$[f_x f_y \; f^2{}_y]$

Response = det M - k (Trace(M))^2
$det([a,b;c,d]) = a*d - b*c$

$$Trace(M) = \sum_{i=1}^{n} M_{ii}$$

Trace is sum of diagonal:

R = $f^2{}_x f^2{}_y - (f_x f_y)^2 - k\,(f^2{}_x + f^2{}_y)^2$

$f^2{}_x = f_x . f_x =$dot product -> shouldn't this be pointwise product (as dot product only exists for vectors)?
0 0 1 0
0 0 1 0
0 0 1 0
0 0 1 0

$f^2_y = f_y f_y =$

0 0 0 0
0 0 0 0
0 0 0 0
0 0 0 0

$f_x f_y =$

0 0 0 0
0 0 0 0
0 0 0 0
0 0 0 0

Response of f^2_x = 4 (sum of pixel values)
$f^2_y =$ 0

In that example:
M= 4 0
0 0

R=4*0-0-k*4^2=-k*16
k=0.04
I read somewhere that k=0.04~0.06 Typo, k was 0.04
R=-0.64 -> edge

R: single number

R>0 -> corner
R<0 -> edge
R~0 -> flat

http://www.songho.ca/dsp/convolution/convolution2d_example.html

c) Edges and corners,
d) Same as b) R = 6,56 so it's a corner
B=
0 0 0 0
1 1 1 0
1 1 1 0
1 1 1 0
$f_x =$

0 0 0 0
0 0 1 0
0 0 1 0
0 0 1 0

$f_y =$

0 0 0 0
1 1 1 0
0 0 0 0
0 0 0 0

$f_x^2 = f_x, \quad f_y^2 = f_y$

$f_x f_y =$

0 0 0 0
0 0 1 0
0 0 0 0
0 0 0 0

$[?][?][?] f = \sqrt{f_x^2 + f_y^2} =$

0 0 0 0
1 1 $\sqrt{2}$ 0
0 0 1 0
0 0 1 0

$M = \sum_{x,y} w(x,y)[f_x^2, f_x f_y; f_x f_y, f_y^2] = [3,1;1,3]$

$R = det(M) - k(trace(M))^2$

$det(M) = 3*3 - 1*1 = 8$

$trace(M) = 3+3 = 6$

$k = 0.04$

$R = 8 - 0.04 * 6^2 = 6.56$

e) $det(M - \lambda I) = 0 \, where \, \lambda \, represents \, the \, eigenvalues$

M (calculated from d)) =
3 1
1 3
I = is the Identity matrix for a 2x2 matrix
1 0
0 1
thus $det(M - \lambda I)$=
3-λ 1
1 3-λ
= 0

(3 - λ)^2 -1 =0 => λ^2 - 6λ + 8 = 0 =>
λ1 = 4
λ2 = 2
Since they are both > 0 it is a corner.

Exercise 3

a) What is an image descriptor?

Image representation to describe a local patch

b) What are the advantages of using histograms as image descriptors? What about quantization (# of bins)?

Histograms: Easy to compute but have a fixed length vector
Quantization (# of Bins): With a higher number of bins, the resolution of the histogram is increased, and it can be more discriminative. However, this is computationally more expensive. With a lower number of bins, the computational cost is reduced, but the resolution is also reduced and the bin values are higher (on average).

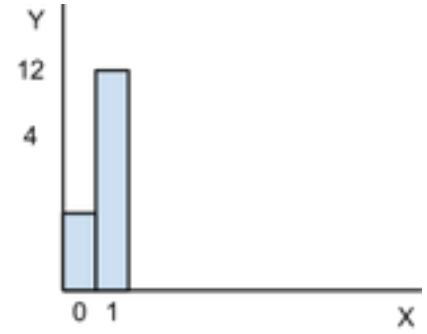
Quantization:
Grids: fast but applicable only with few dimensions
Clustering: slower but can quantize data in higher dimensions

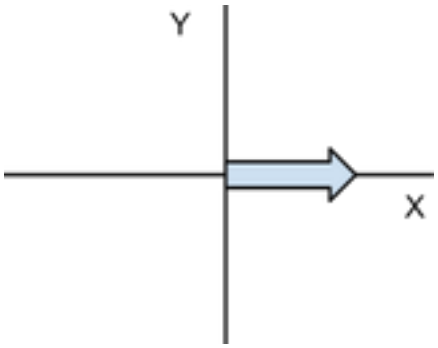
c) What kind of image structures are descriptors made of?

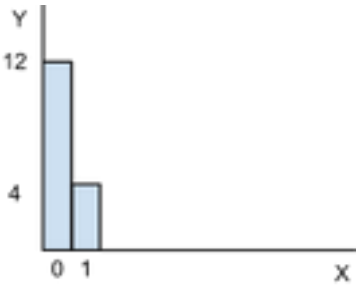
Pixel intensities, oriented gradients and color

d) Compute the histogram of oriented gradients and pixel values for patch A given in Figure 1.

compute fx and fy and sum the pixels to check the orientation
Histogram for A

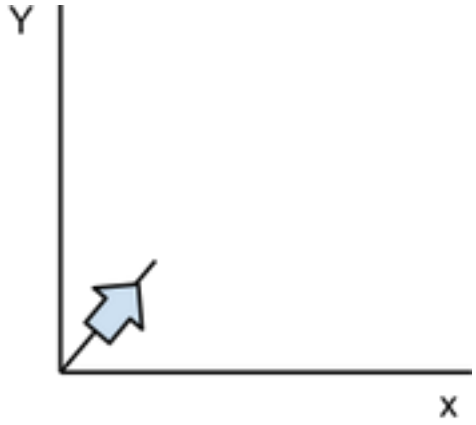






Histogram of gradient

e)
fxfx =
0 0 0 0
0 0 1 0
0 0 0 0
0 0 0 0



f) What is the SIFT descriptor?

The SIFT descriptor describes the local shape of a region/patch by using gradient orientation histograms.

g) Is the SIFT descriptor invariant under a change in (in-plane) rotation of the object? Please explain.

We can make it invariant to rotation by normalizing on the maximum gradient orientation.

h) What is a color SIFT descriptor? Give an example of robust color SIFT descriptor.

A color SIFT descriptor applies the SIFT descriptor to color channels.
Examples: rgSIFT, HSV-SIFT, OpponentSIFT.
A robust one is RGB-SIFT.

Exercise 4

a) What is the difference between dense and point sampling?

In dense sampling we take every possible patch as a sample whereas in point sampling we only take the interest points as samples.

b) What are the basic steps of the bag-of-feature approach?

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies of “visual words”

c) What are visual words and how is the visual vocabulary computed?

Visual Words can be represented by small parts of an image which carry some kind of information related to the features (such as the color, shape or texture), or changes occurring in the pixels such as the filtering, low-level feature descriptors.
They are computed by counting the number of appearances (frequency).

d) What are spatial pyramids and why are they useful?

Is an extension of a bag of features and a locally orderless representation at several levels of resolution. It is useful for discriminating between histograms for different resolutions.

e) Using the back-of-features approach with SVM for object recognition, do you expect that certain objects may be confused during recognition? Give examples.

Some object may have similar descriptions and get misclassified. (dogs classified as cats, planes classified as birds viceversa)

f) Do you think that context is important for object recognition? Can you give an example of certain objects?

Object that can be found indoor and outdoor may vary.

Exercise session 5 (05.12.2012)

Exercise 1

a.
s(x,y) =
5 7 9
11 17 24
13 22 35

There was an error in the third row; the correct number should be 22, not 24 (3 + 13 + 17 - 11)
b.
S(A) + S(D) - S(B) - S(C) = 5+35-9-13 = 18

(note: above and left of the corner A is 5 2x2 subrectangle)
c.
O(I) (don't know whether it is 1 or i), I assume it's "i" (no of pixels in the image)
I would say it's O(1), since the above computations are done in constant time, regardless of the dimension of the considered patch. Agreed.
d.

s(x,y) =
1 2 3 4
2 4 6 8
2 4 6 8
2 4 6 8

e.
Horizontal and vertical edges.
f.
8-0 = 8 (note: above and left values are 0 for corner A)

This is the matrix. Follow the White Rabbit...

Exercise 2

a. What is the basic pipeline for window-based object detection?

1) Build an object model and learn to classify
2) Generate candidates in the new image
3) Score each candidate

b. What are the advantages of a sliding-window approach?

Advantages: Simple, yields good results
Disadvantages: Expensive in computational time

c. Given an image of 256x256, how many windows are required to detect objects for 8 different orientations and 6 scales?

256x256x8x6 = 3,145,728

d. Assume that for a strong classifier (e.g. non-linear SVM), the time for window classification is about 0.01 seconds. How many hours does it take to detect objects in 10,000 images?

We take the result from C multiplied with the time for 1 window (0.01) and with the number of images and divide it by the 3600 seconds: **87381 hours**

e. How can one reduce the number of bounding boxes for detection? Now, how long does it take to detect objects in 10,000 images?

We use image segmentation, and each window will become a segment. This way we will have an average of 1500 windows and a recall of 98%.
Alternatively we could jump locations or reuse calculations if they overlap.
For the 10k collection it will take approx. 41 hours.

Exercise 3

a. What is an action unit?

AU's are systematically defined labels to describe movements of individual (or groups of) facial muscles and the resultant appearance. FACS = acronym for Facial Action Coding System.

b. What are the six basic emotions of humans?

Happiness, Sadness, Disgust, Fear, Surprise, Anger

c. What are the facial characteristics of a happy person?

Upward movement of the mouth corners which correspond to AU12. Duchenne markers, cheeks up, etc.
In terms of anatomy Zygomatius Major contracts and moves the lip corners upward.

d. Considering Figure 3, which AU's indicate positive emotions/expressions?

Positives AU - 12, 6 (1+2 >> surprise)

e. Which ones of them indicate negative emotions/expressions?

Negative AU - rest of them and (1+2 >> surprise)

f. Which of the action unit or action unit combination can describe sadness expression?

1, 4, 15

g. Which of the action unit or action unit combination can describe a smile?

Smile = AU12 or Duchenne smile 12+6

h. Which features are important to recognize a real enjoyment smile?

Dynamics (speed, acceleration, amplitude) of lip corners and eyelids movements (the orbicularis oculi muscle)