The Canonical Correlations of Color Images and their use in Inverse Problems

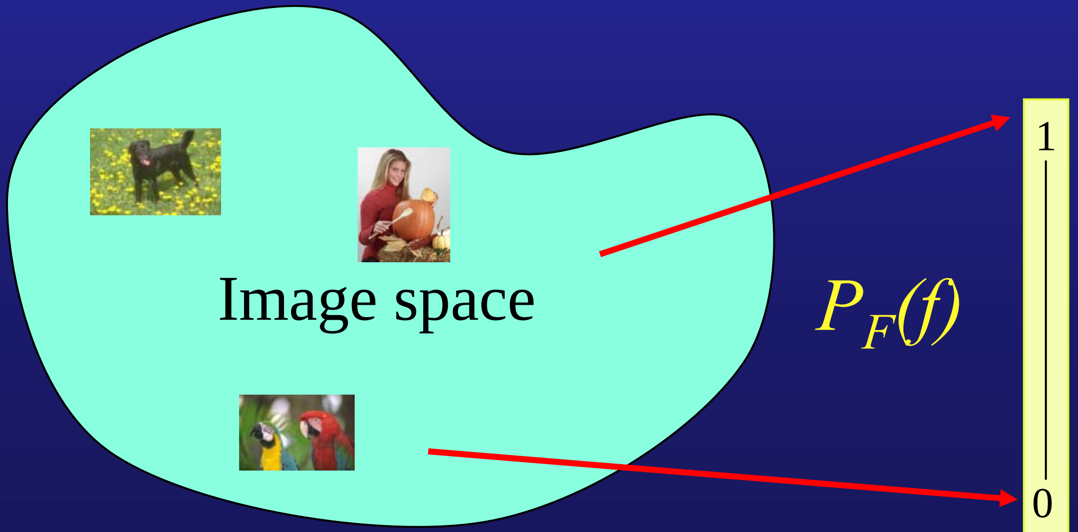
Yacov Hel-Or
School of Computer Science
The Interdisciplinaty Center

The Image Prior

- A color image is typically represented by three bands:

$$\boldsymbol{f}(x,y) = [\ R(x,y)\ G(x,y)\ B(x,y)\ ]^T$$

- Independent of the representation, a prior statistical distribution over natural images is required in many applications:



Does the HVS use an Image prior?

Image Prior and Inverse Problems

- An ***Inverse problem*** of color images aims at reconstructing an image $\mathbf{f}(x,y)$ from its degraded version $\mathbf{m}(x,y)$:

$$\mathbf{m}(x,y) = \underbrace{D[\mathbf{f}(x,y)]}_{\text{Degradation Model}}$$

- The degradation operation is non-invertible or ill-posed!
- Examples:
  - Image demosaicing.
  - Image Scaling.
  - Image Sharpening.
  - Image Denoising.

- A possible solution using the *Maximum a Posteriori* (MAP) estimator:

$$\hat{f} = \arg\max_f P_{F|M}(f \mid m)$$

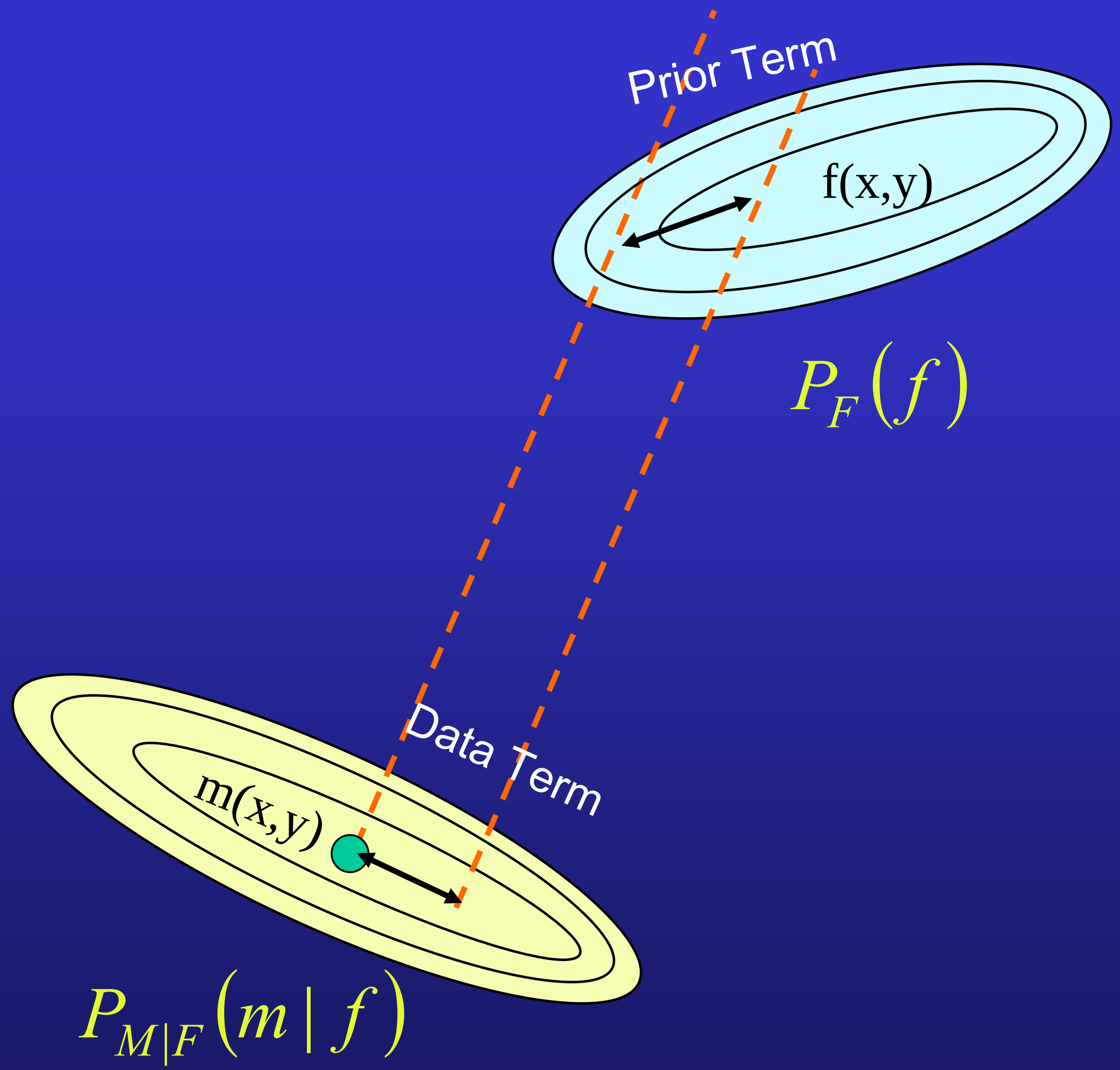
- Using Bayes conditional rule:

$$\hat{f} = \arg\max_f P_{M|F}(m \mid f) P_F(f)$$

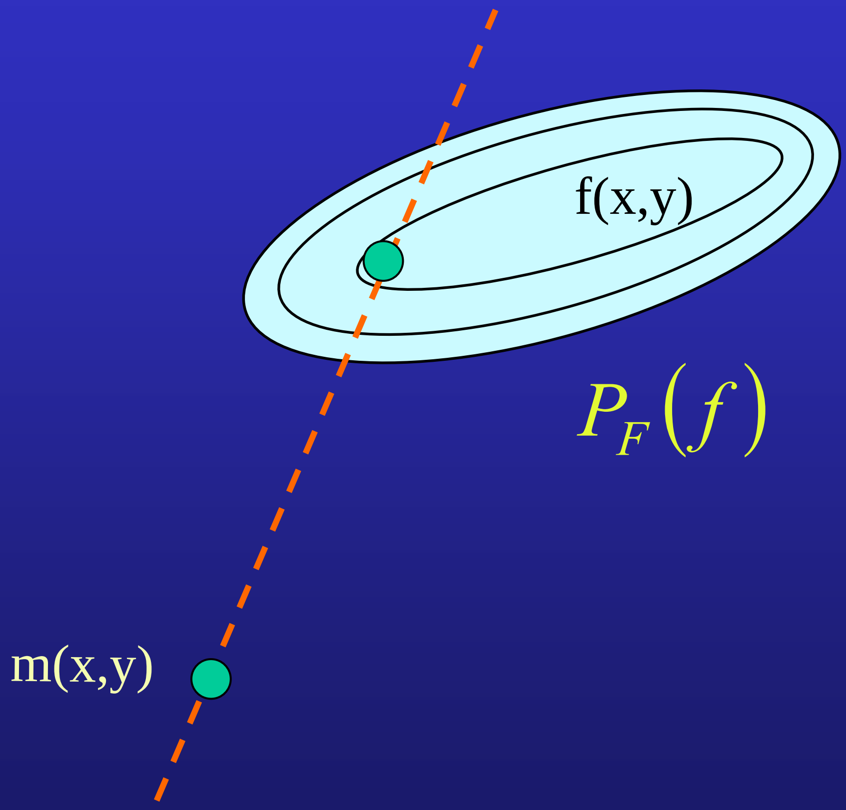
$$= \arg\max_f \underbrace{\log P_{M|F}(m \mid f)}_{\text{Degradation Model (data term)}} + \underbrace{\log P_F(f)}_{\text{Image Prior Model (prior term)}}$$

Degradation Model
(data term)

Image Prior Model
(prior term)


Prior Term
f(x,y)
$P_F(f)$
Data Term
m(x,y)
$P_{M|F}(m \mid f)$

- If the degradation model is noise free, the "data term" has a ridge distribution and becomes a constraint:



- The “data term” $P_{M|F}$ is derived from the degradation process and is (relatively) easy to model (Gaussian Noise, noise free).

- The “prior term” P_F defines a prior over natural color images:
  - Defined over a huge dim. space (3E6 for 1Kx1K color image)
  - Known to be non Gaussian.
  - *Very* complicated to model.
  - Crucial for any reconstruction method.

Main Goal:

Modeling a (useful) prior distribution of natural color images

“All statistical models are wrong, but only some are useful”

Quoted from: *Statistics of Images* .. by Mark. L. Green.

Towards Useful Priors:
Dimensionality Reduction

- Due to the dimensionality of P_f, modeling the entire joint distribution is impractical.
- In order to build a *useful* model we must reduce the dimensionality of the problem.

- Common approaches for image modeling use 2 types of reductions:
  - Reduction in the Spatial domain.
  - Projection onto *informative* subspaces.

Reduction in the Spatial Domain

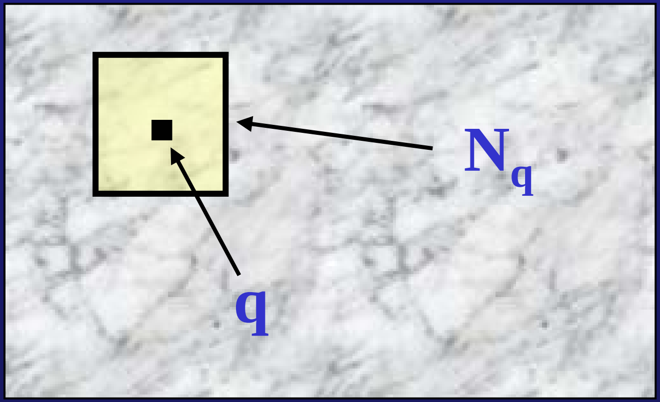
- A reasonable assumption: A natural image can be viewed as a realization of a ***Markov Random Filed***:
  1. A large enough neighborhood of an image pixel completely characterizes its p.d.f.:

  $$P\left(q \mid N_q\right) = P\left(q \mid p, p \neq q\right)$$

  2. This p.d.f. is similar for all pixels (the homogeneity property of images).

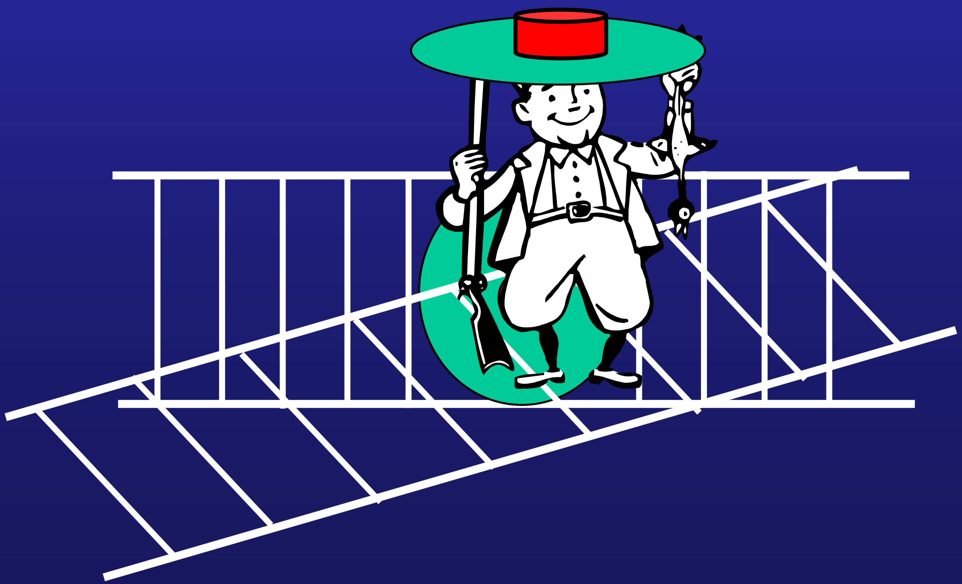
- We have to model only the distribution of local contexts:

$$P\left(q, N_q\right)$$



Projection onto Informative Subspaces

- Further reduction can be achieved by modeling only marginal distributions over subspaces of the context space.
- Subspaces should be chosen such that "*informative*" information will not be lost.
- A crucial problem: what are "*informative*" subspaces?



Informative Subspaces

- Informative subspaces are task driven.
- If our task is to predict **y** form ***x*** (and visa versa) we should choose subspaces in which the two variables are most correlated.

The Canonical Correlation Analysis (CCA)

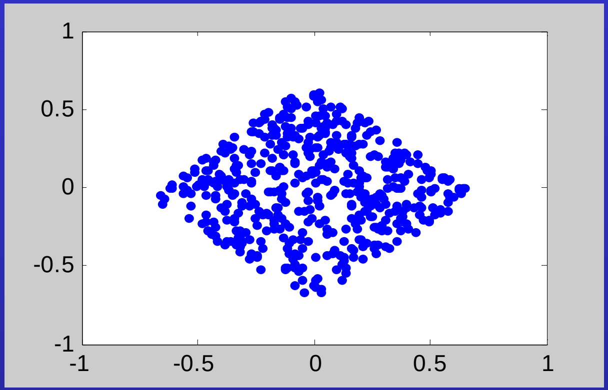
finds such subspaces.

The Canonical Correlation Analysis (CCA)

- Assume two multidimensional random variables: **x** and **y** .
- We are looking for two projection vectors w_x and w_y such that the correlation between $x' = \mathbf{x}^T \mathbf{w}_x$ and $y' = \mathbf{y}^T \mathbf{w}_y$ is maximized:

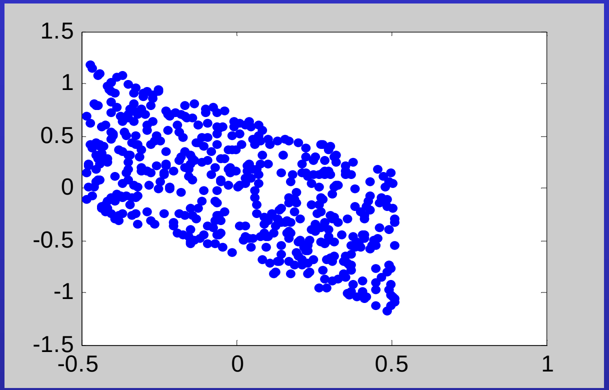
$$\rho(w_x, w_y) = \frac{E\{x'y'\}}{E\{x'^2\}E\{y'^2\}}$$

Simple Example



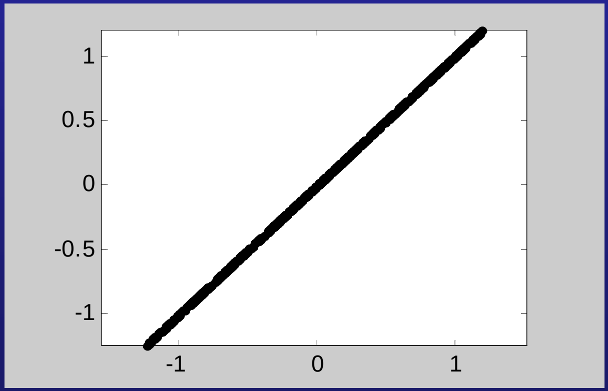
$\mathbf{x}=(x_1,x_2)$

$x_1 - x_2 = 2y_1 + y_2$



$\mathbf{y}=(y_1,y_2)$

x’



y’

$x' = (1, -1)\,\mathbf{x}$

$y' = (2, 1)\,\mathbf{y}$

Canonical Correlations

The Canonical Correlation Analysis (CCA) Hoteling 1936

- The solution for w_x and w_y satisfies the eigenvalue equations:

$$C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy} w_y = \rho^2 w_y$$

$$C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} w_x = \rho^2 w_x$$

where:

$$C_{xx} = \frac{1}{n} \sum_i \mathbf{x}_i \mathbf{x}_i^t \qquad C_{yy} = \frac{1}{n} \sum_i \mathbf{y}_i \mathbf{y}_i^t \qquad C_{xy} = \frac{1}{n} \sum_i \mathbf{x}_i \mathbf{y}_i^t$$

- The CCA characteristics:
  - $w_{x,1}$ and $w_{y,1}$ corresponding to the greatest eigenvalue ρ_1^2 define the directions with the maximal correlation ρ_1.
  - The subsequent 2 eigenvectors are the second best directions, and so forth.
  - The set of eigenvectors are the ***CC basis vectors***.
  - The corresponding eigenvalues are the ***Canonical Correlations***.
  - The CCA basis vectors also decorrelate off-diagonal terms, i.e. C_{xx} C_{yy} and C_{xy} are diagonal in the new basis
  - If x and y are Gaussians CCA maximizes also their mutual information.

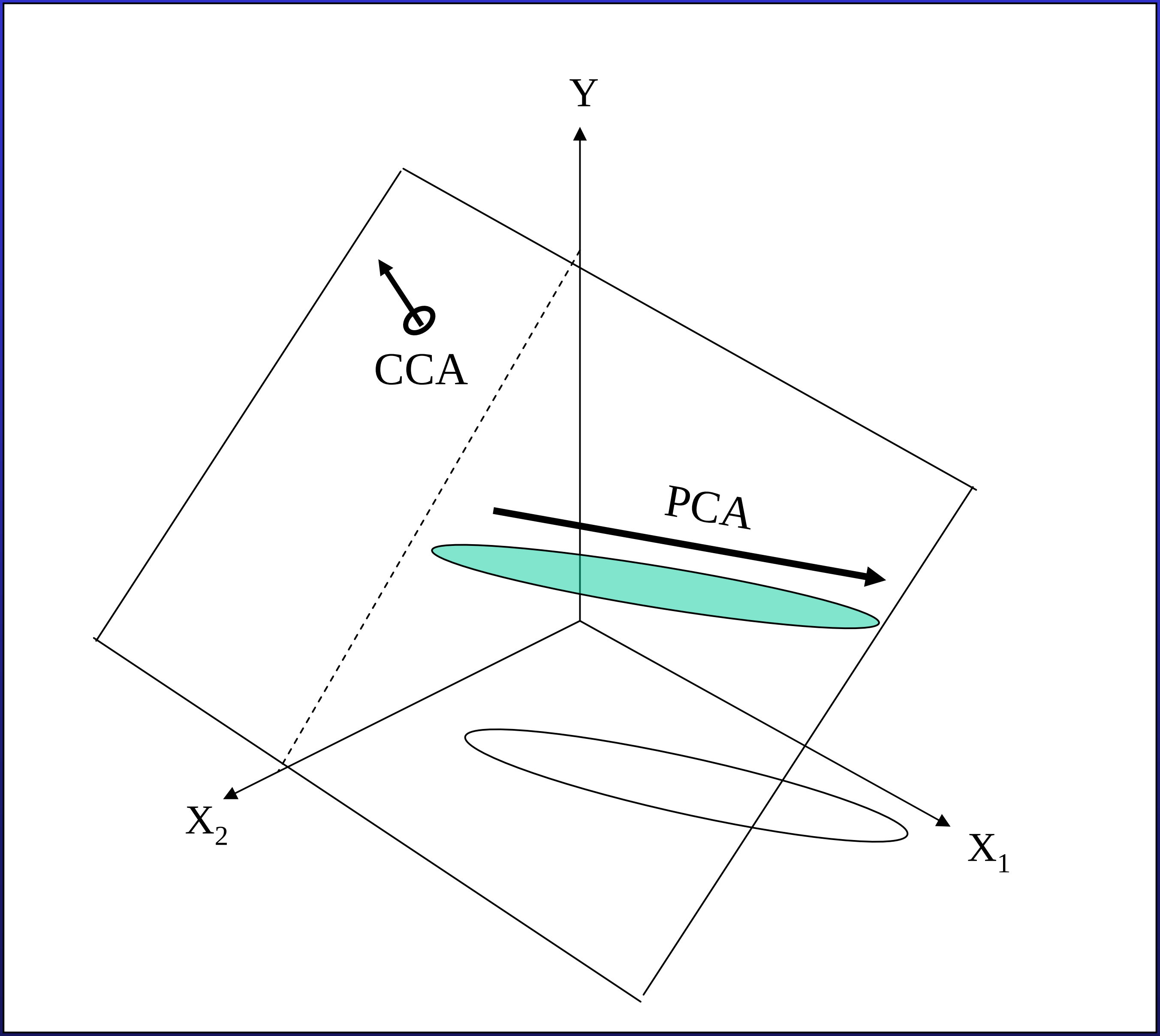
CCA V.S. PCA

	CCA	PCA
Variables	distinct entities	augmented
Mutual Correlations	Maximizes	Minimizes
Affine Trans.	independent	dependent
Consideration	*between* classes	*between* and *within* classes

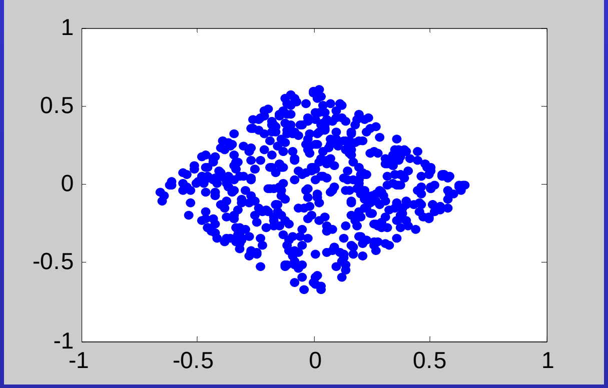
Simple Example

$y=1-x_2+n$

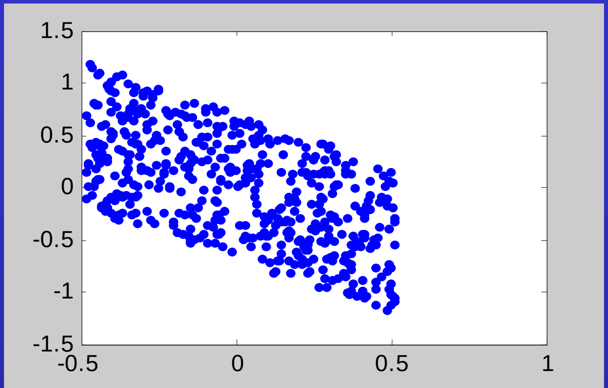
Due to the *within* correlations of *x*, PCA fails to provide useful information.



Previous Example

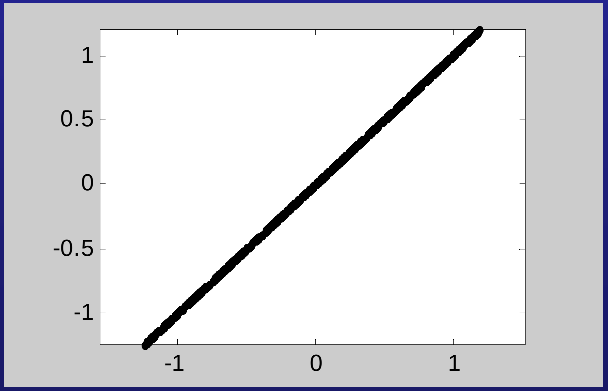


$\mathbf{x}=(x_1,x_2)$

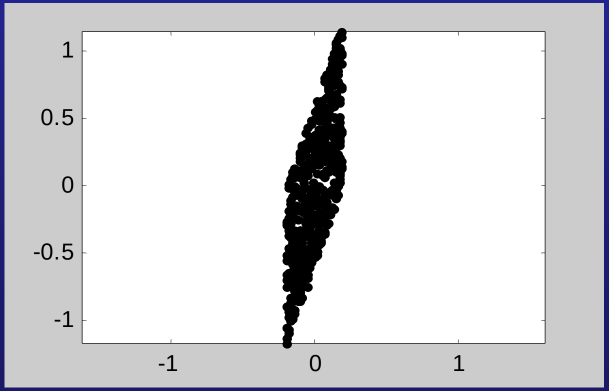


$\mathbf{y}=(y_1,y_2)$

$$x_1 - x_2 = 2y_1 + y_2$$



Canonical Correlations



Principal Components

The CC of Color Images

- In the following we consider natural color image.

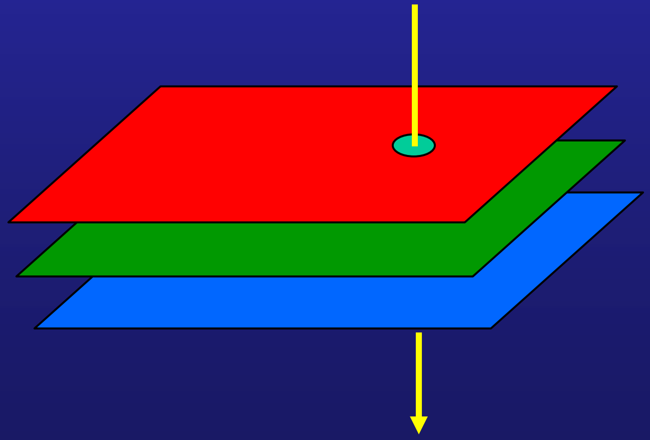
- All values are presented in log(RGB) space:

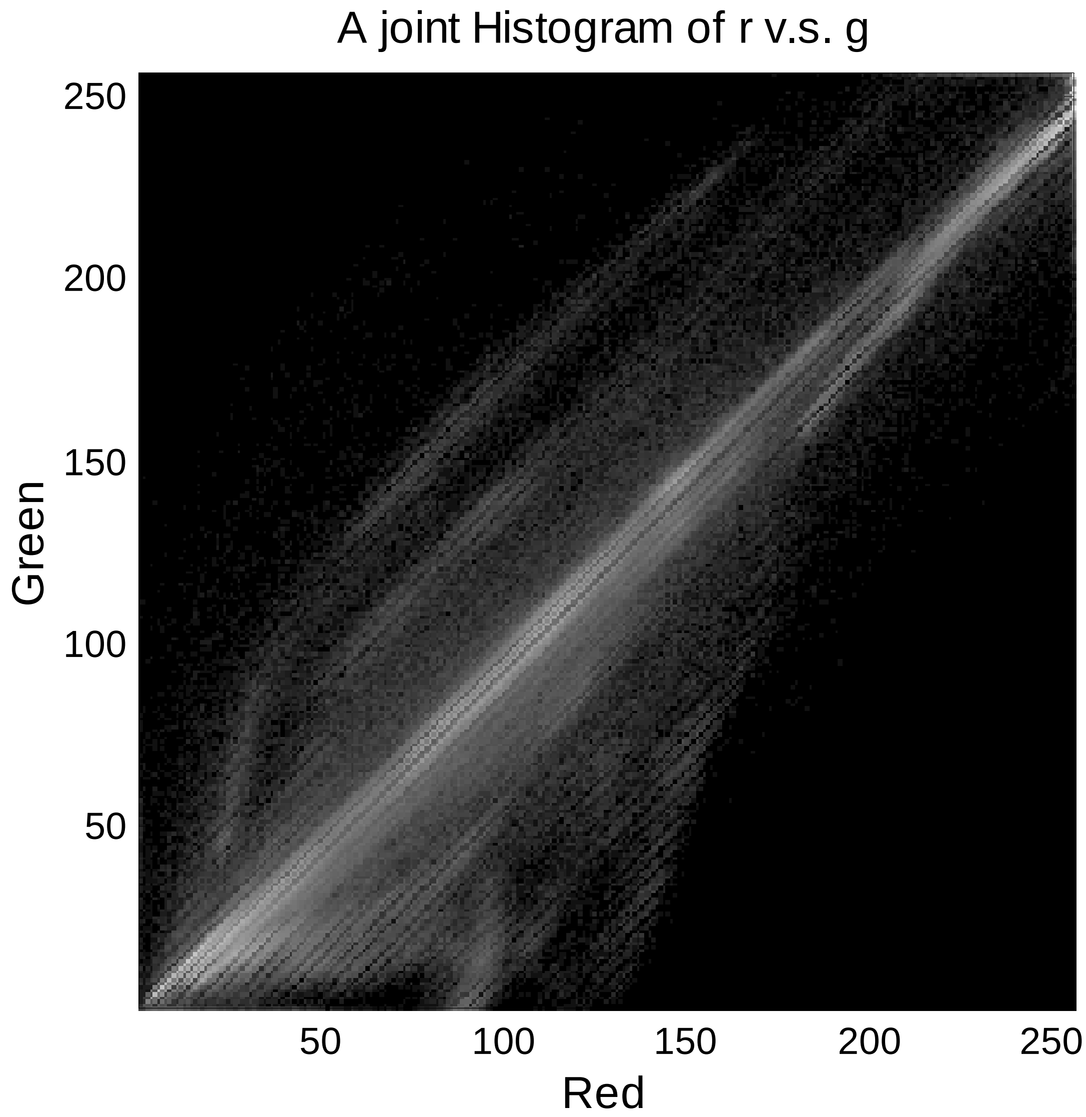
$$\mathbf{f}(x,y) = \log\begin{pmatrix} R(x,y) \\ G(x,y) \\ B(x,y) \end{pmatrix}$$

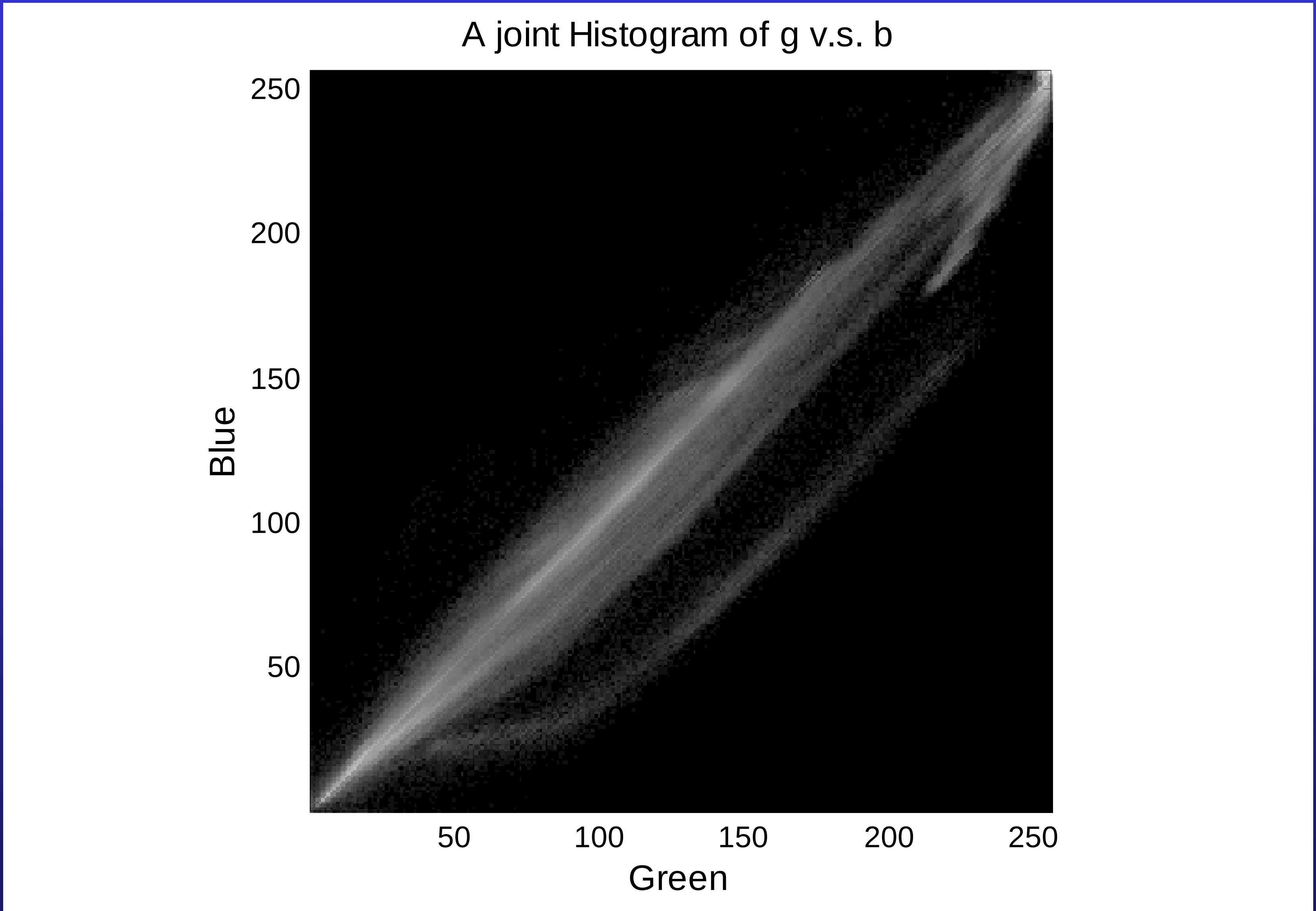
A Special Case of Marginal: Pure Spectral

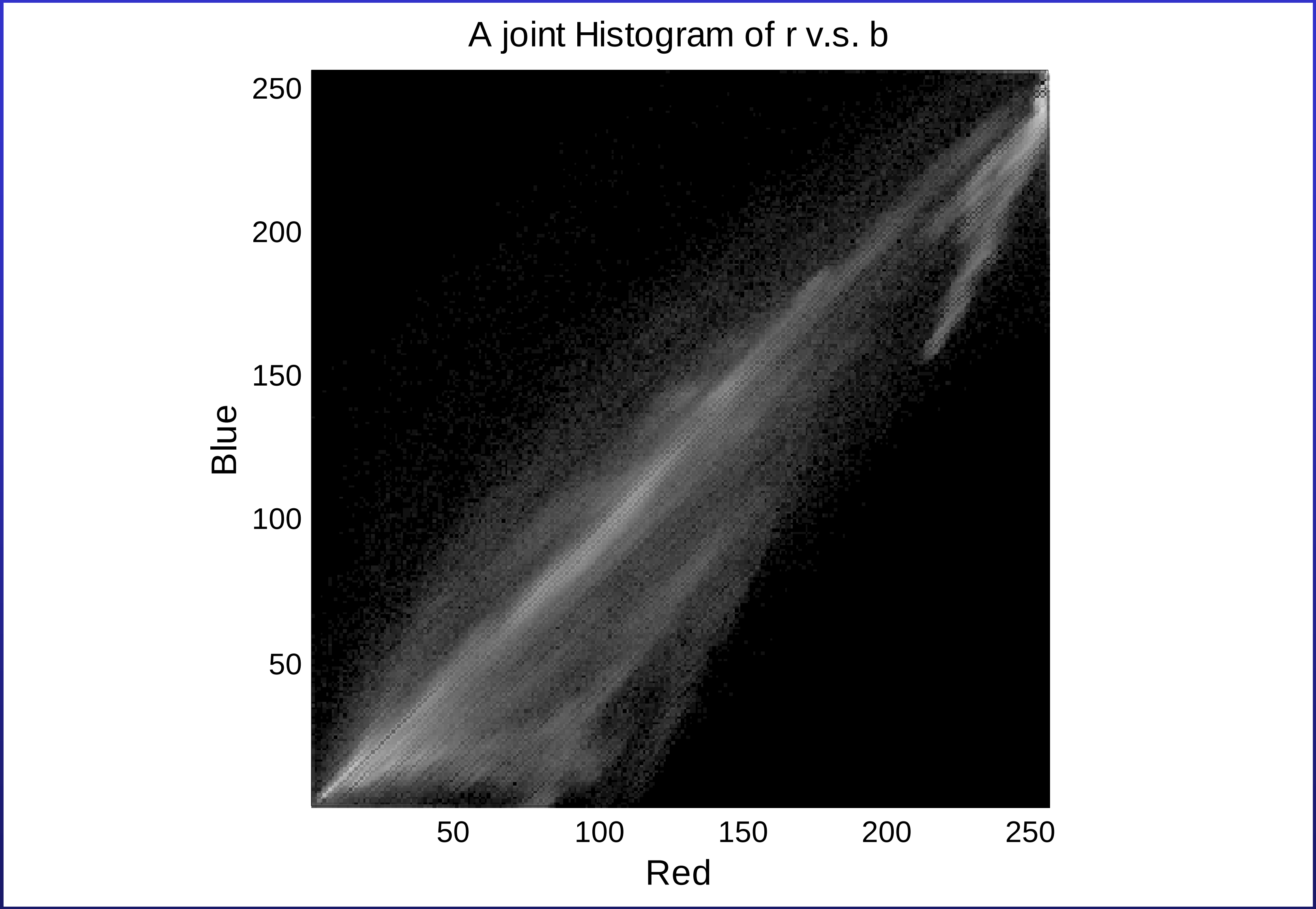
- The following joint histograms show the marginal p.d.f. of the image in the pure spectral domain:

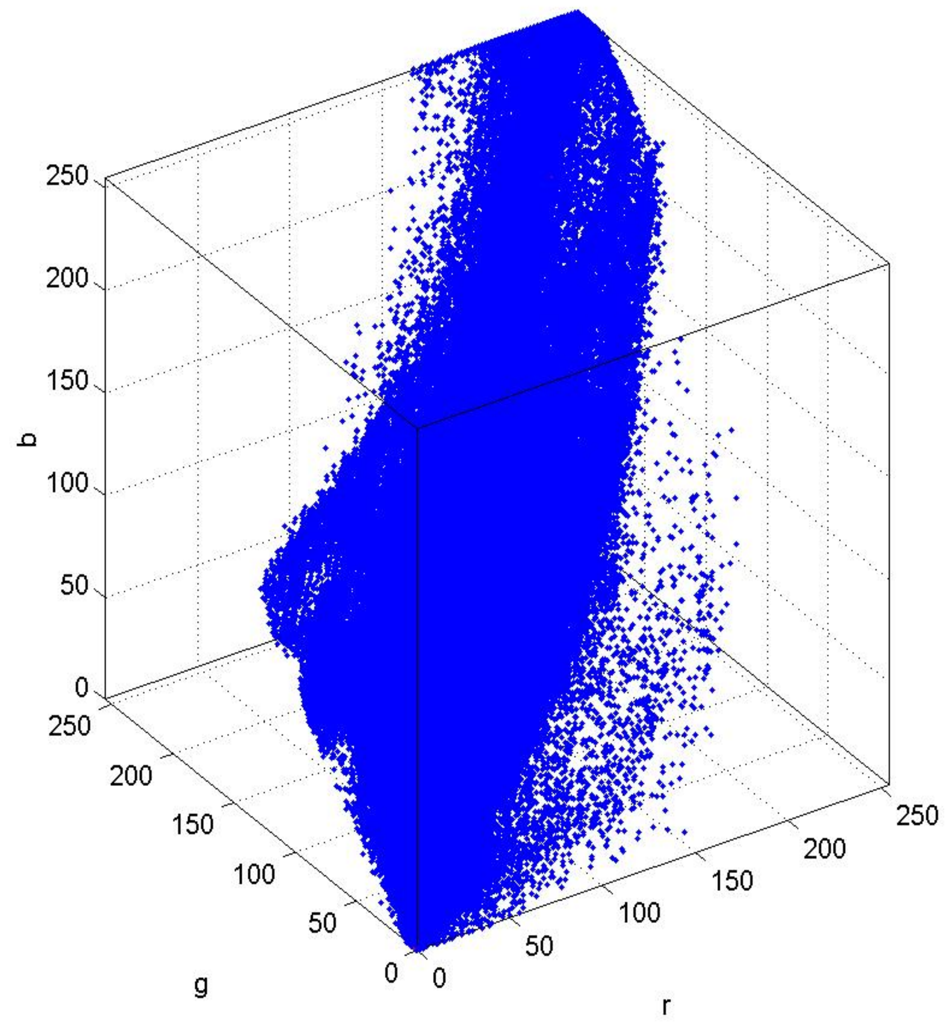
$$P(r,g,b) \propto \#\left(R(x,y)=r, G(x,y)=g, B(x,y)=b\right)$$




A joint Histogram of r v.s. g
250
200
150
100
50
Green
50
100
150
200
250
Red


A joint Histogram of g v.s. b
Blue
250
200
150
100
50
50
100
150
200
250
Green


A joint Histogram of r v.s. b
Blue
250
200
150
100
50
50
100
150
200
250
Red



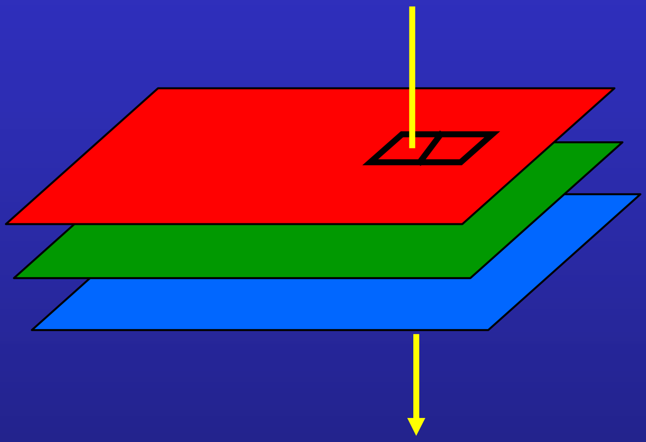
b
250
200
150
100
50
0
g
250
200
150
100
50
0
r
0
50
100
150
200
250

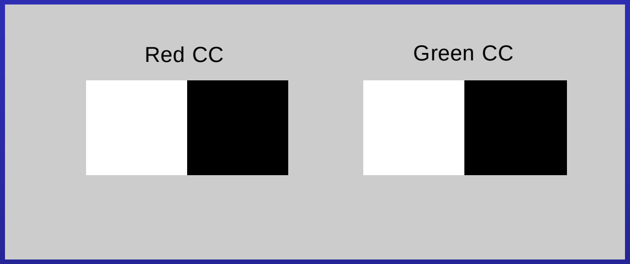
- Observations:
  - The spectral components are correlated.
  - There are diagonal line structures in the histograms.
  - The histogram shapes are not stable over different images.

- Question: Do we loose information if we model the image prior over the pure spectral domain?

The CCA of 1x2 neighborhoods

- Applying the CCA over (R,G,B) where each variable is a 1x2 neighborhood gives the following results:

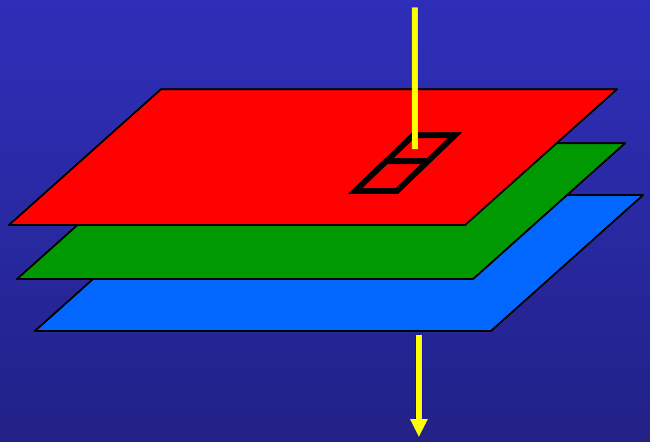


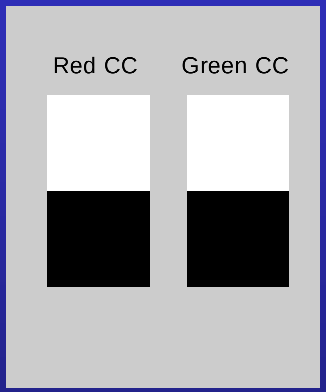


- The CC basis is composed of x-derivatives
- Similar results for each color pair.

The CCA of 2x1 neighborhoods

- Applying the CCA over (R,G,B) where each variable is a 2x1 neighborhood gives the following results:

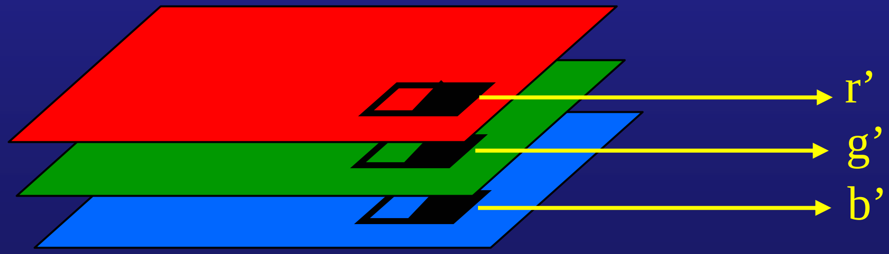


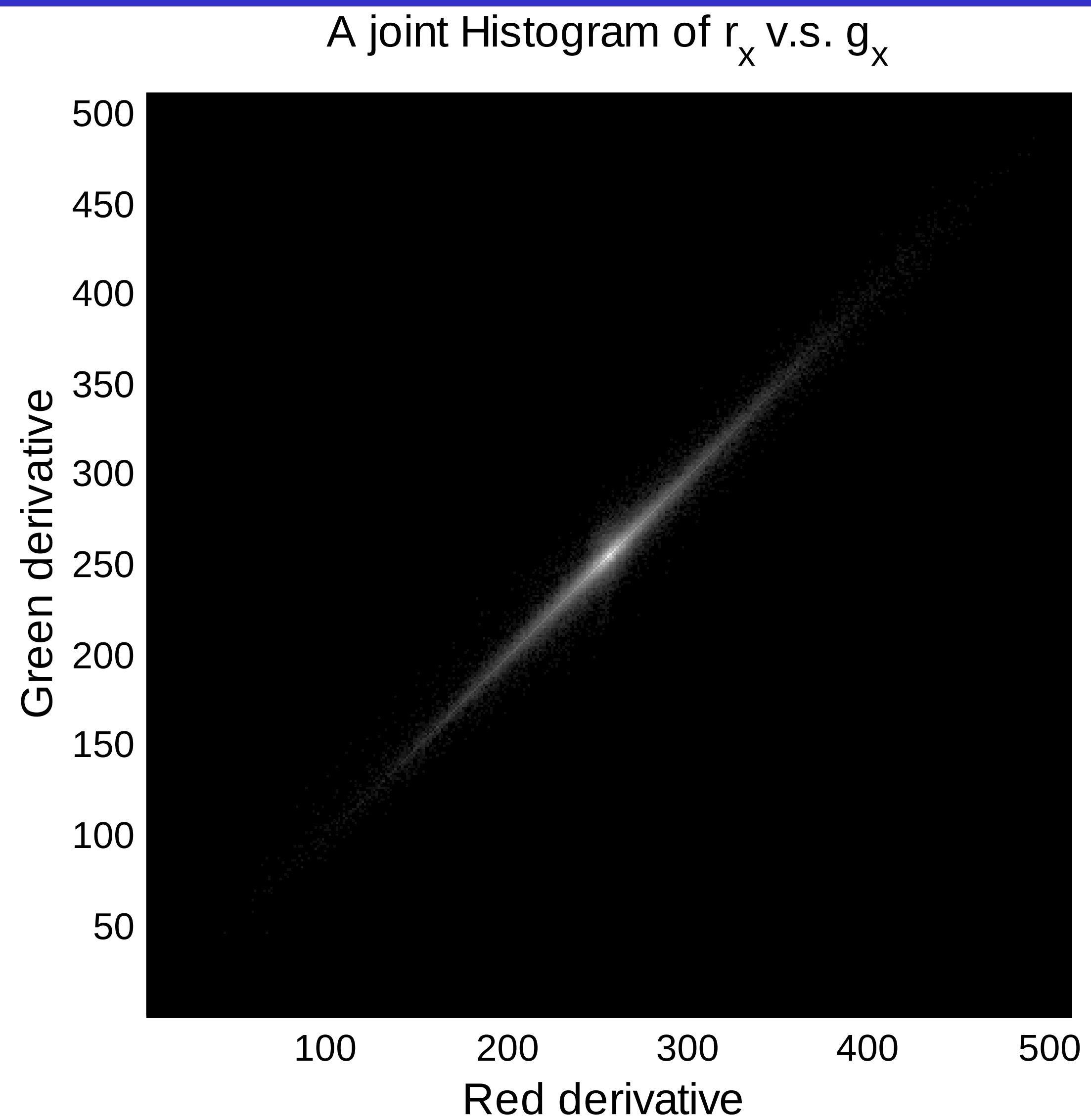


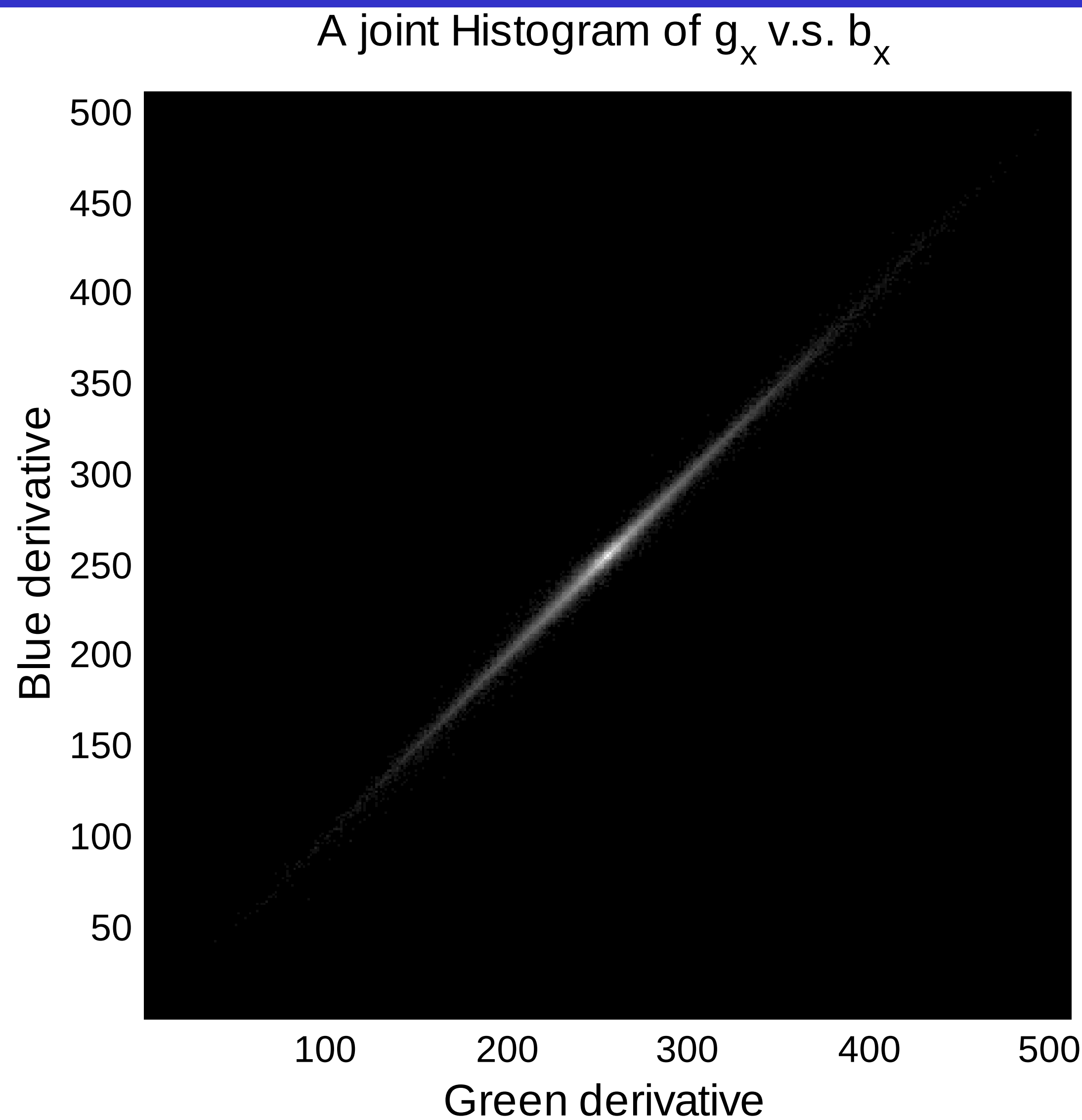
- The CC basis is composed of y-derivatives
- Similar results for each color pair.

- The following joint histograms show the marginal p.d.f. along the first CCA direction, for 1x2 neighborhoods:

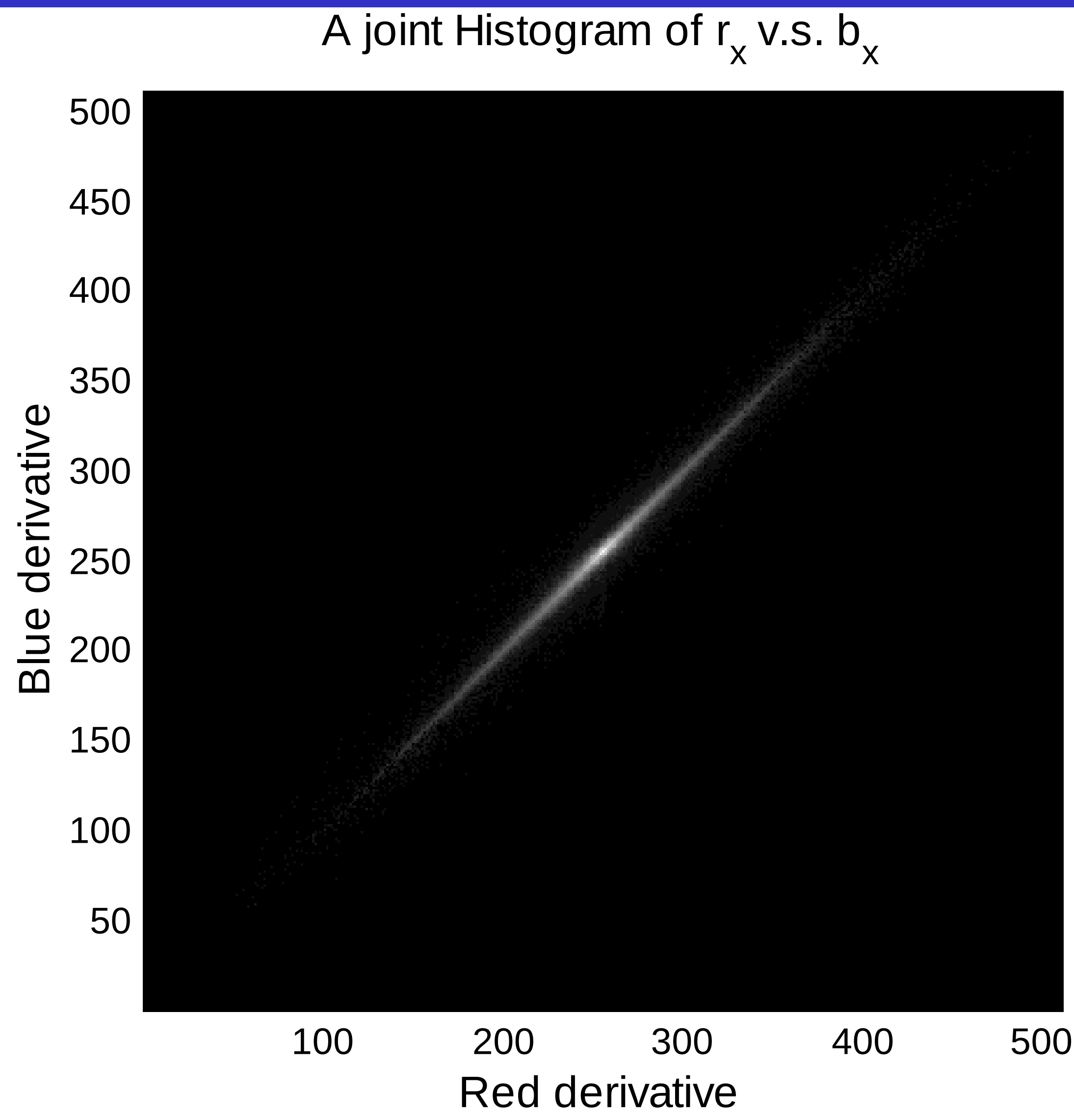
$$P(R',G',B') \propto$$
$$\propto H\left(R^T(x,y)w_R, G^T(x,y)w_G, B^T(x,y)w_B\right)$$



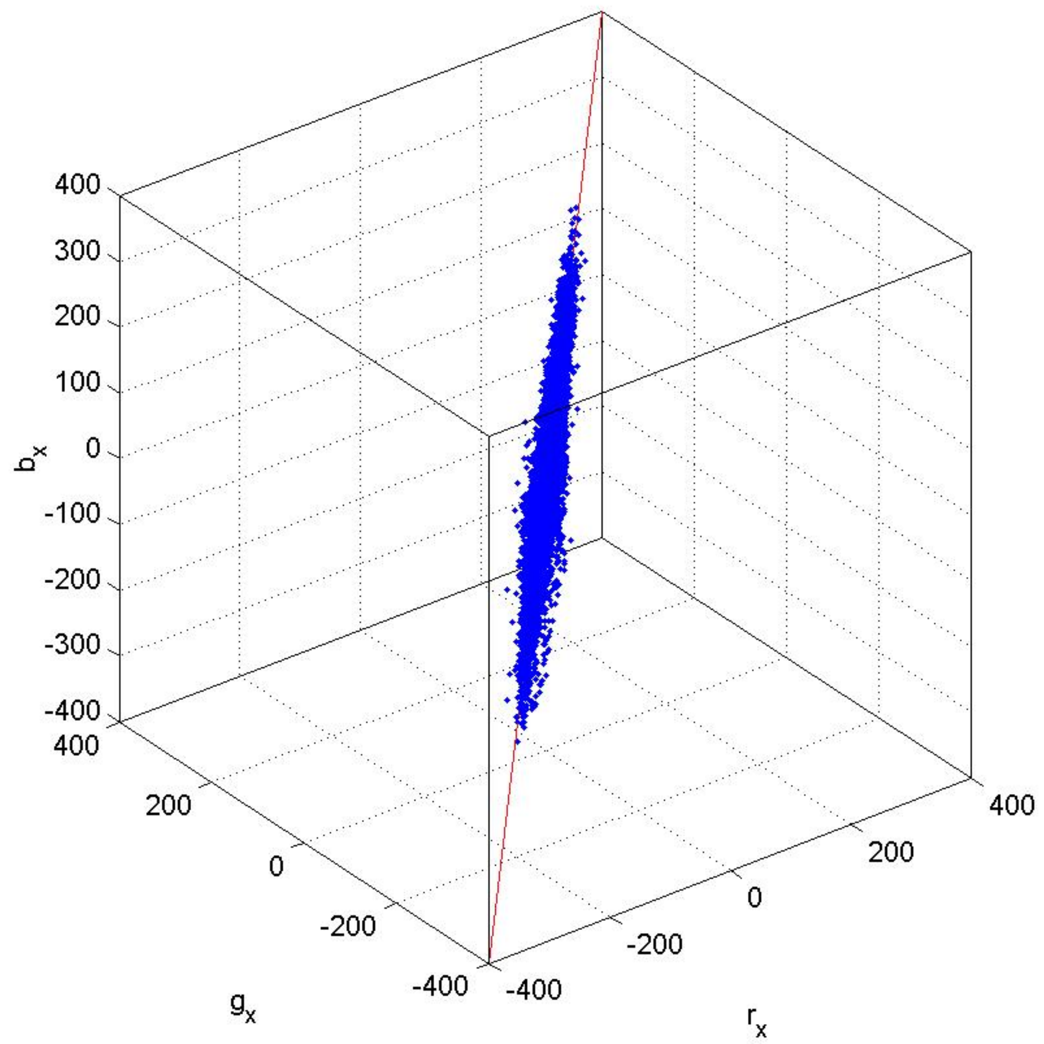

A joint Histogram of r_x v.s. g_x
500
450
400
350
300
250
200
150
100
50
Green derivative
100
200
300
400
500
Red derivative



A joint Histogram of g_x v.s. b_x
500
450
400
350
300
250
200
150
100
50
Blue derivative
100
200
300
400
500
Green derivative



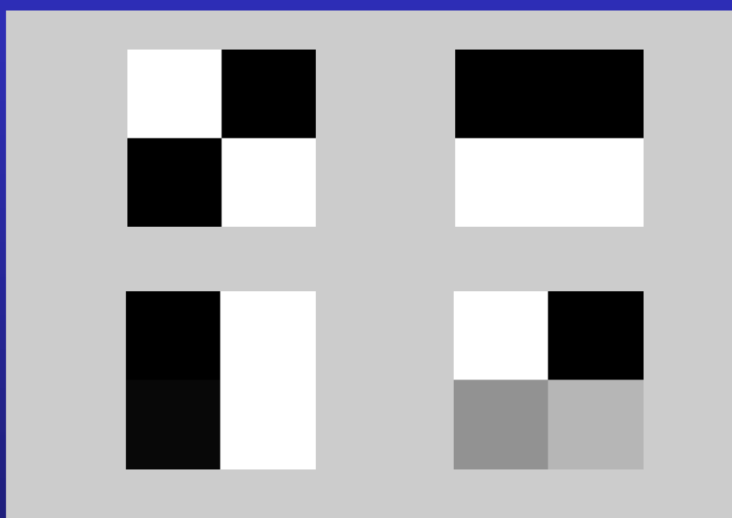
A joint Histogram of r_x v.s. b_x
500
450
400
350
300
250
200
150
100
50
Blue derivative
100
200
300
400
500
Red derivative



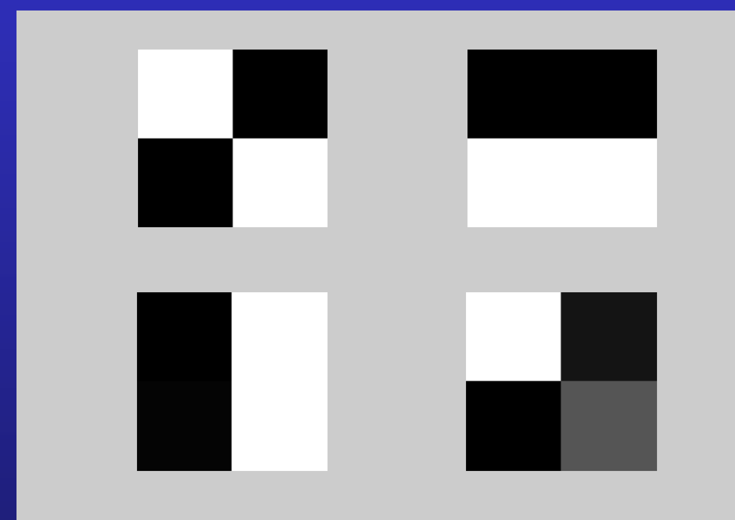
400
300
200
100
0
-100
-200
-300
-400
b_x
400
200
0
-200
-400
g_x
-400
-200
0
200
400
r_x

The CCA of 2x2 neighborhoods

- Applying the CCA over (R,G,B) where each variable is a 2x2 neighborhood gives the following results:

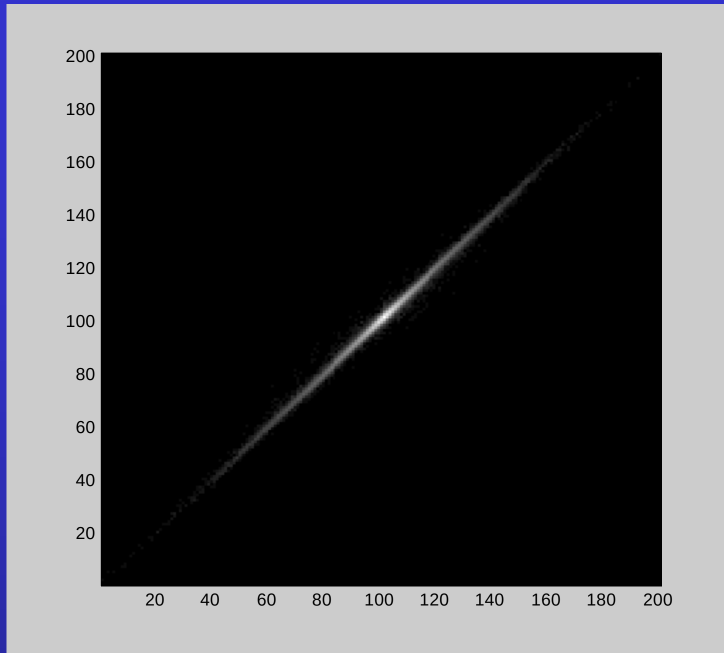


The 4 CC vectors of the Red plane

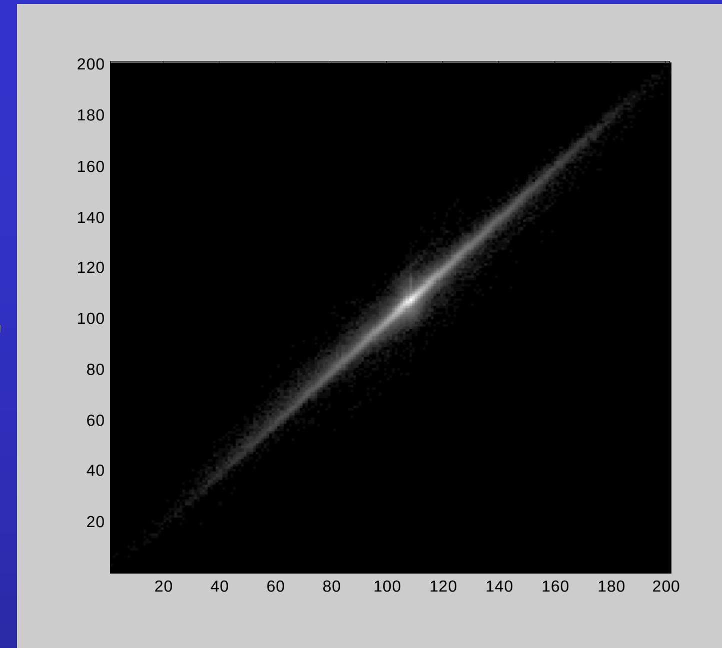


The 4 CC vectors of the Green plane

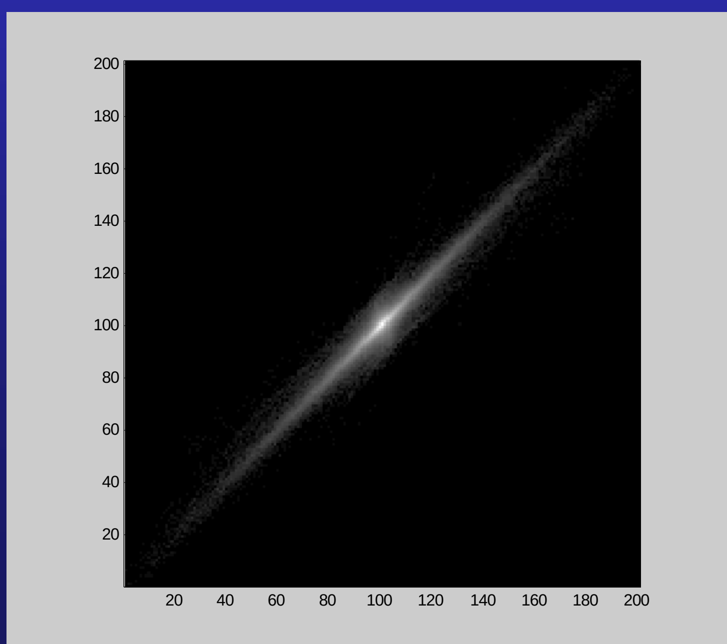
1st CC



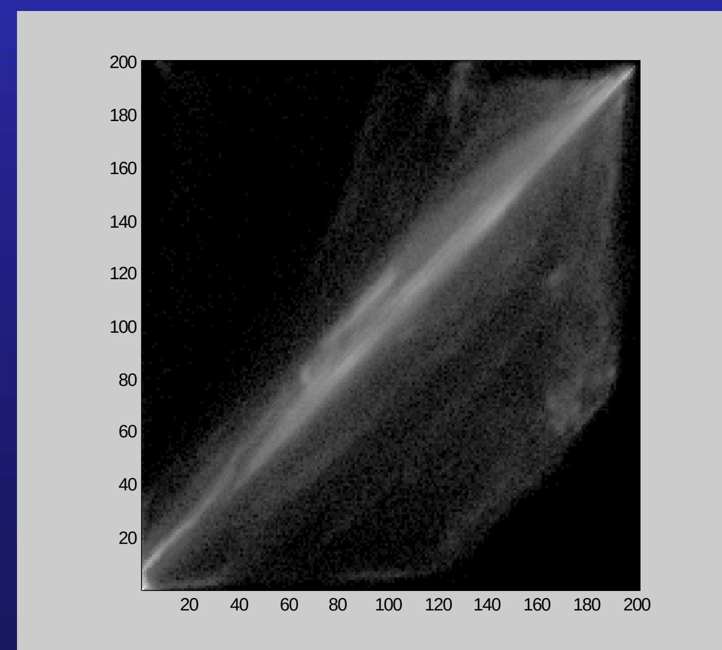
2nd CC

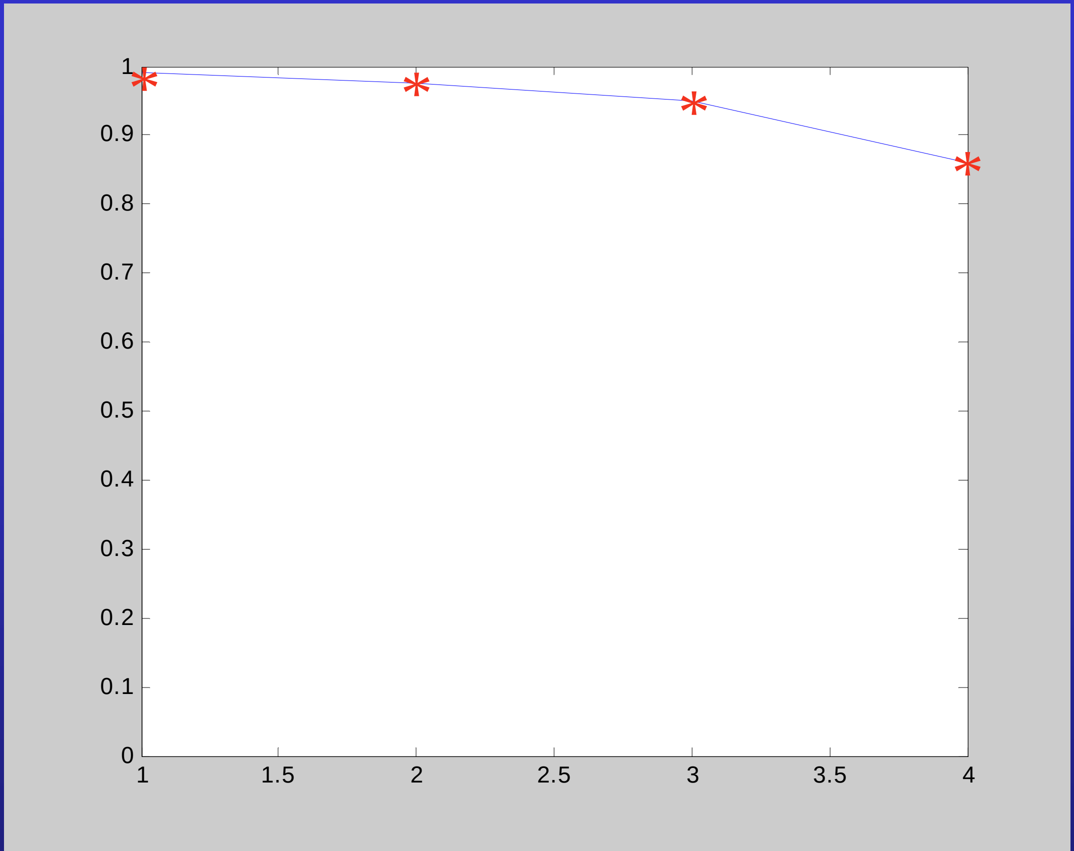


3rd CC

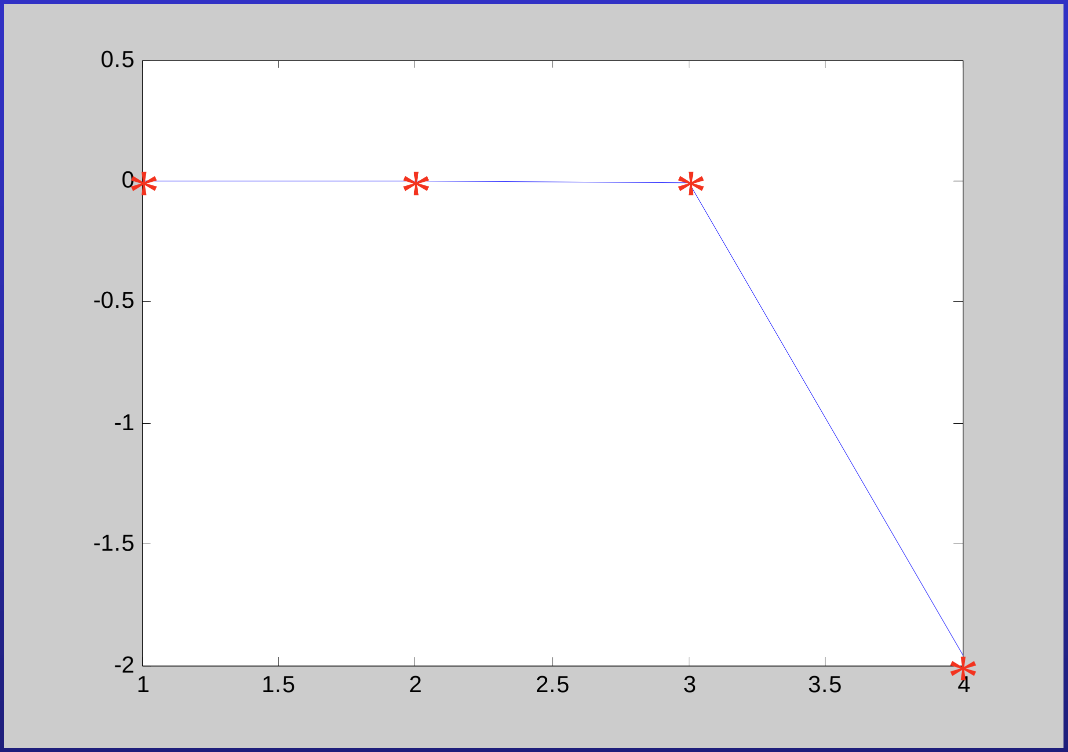


4th CC





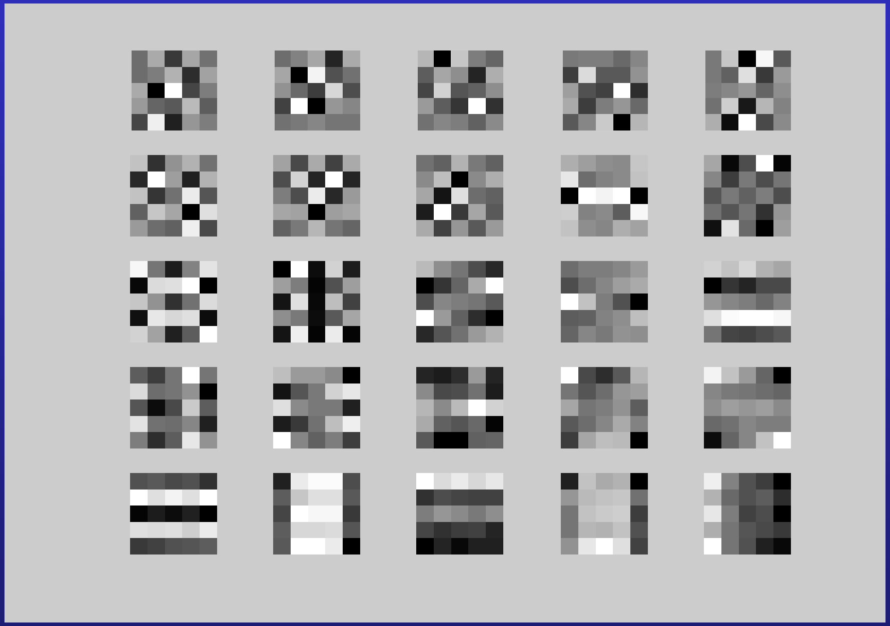
The 4 CC values for the CC vectors



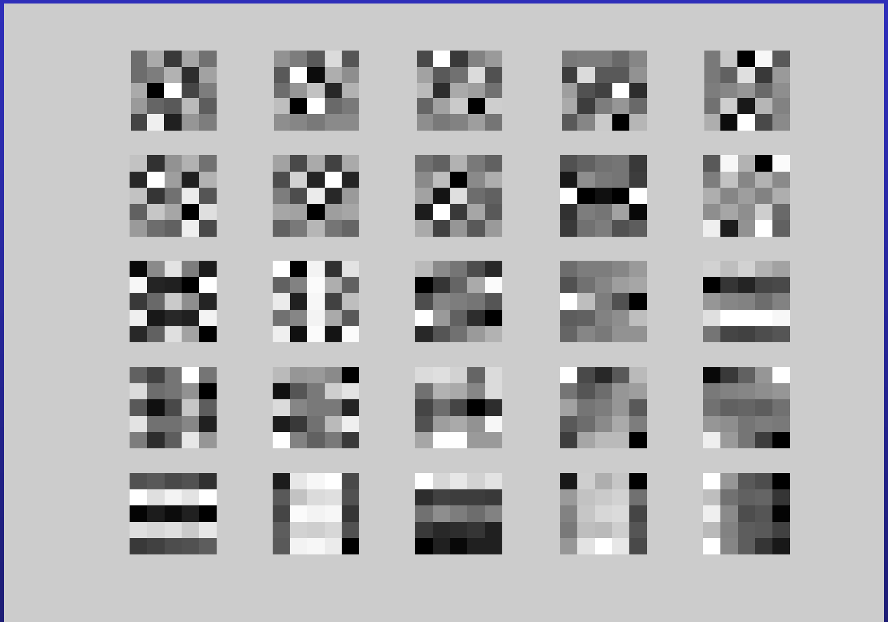
The sum of components of CC vectors

The CCA of 5x5 neighborhoods

- Applying the CCA over (R,G,B) where each variable is a 5x5 neighborhood gives the following results:

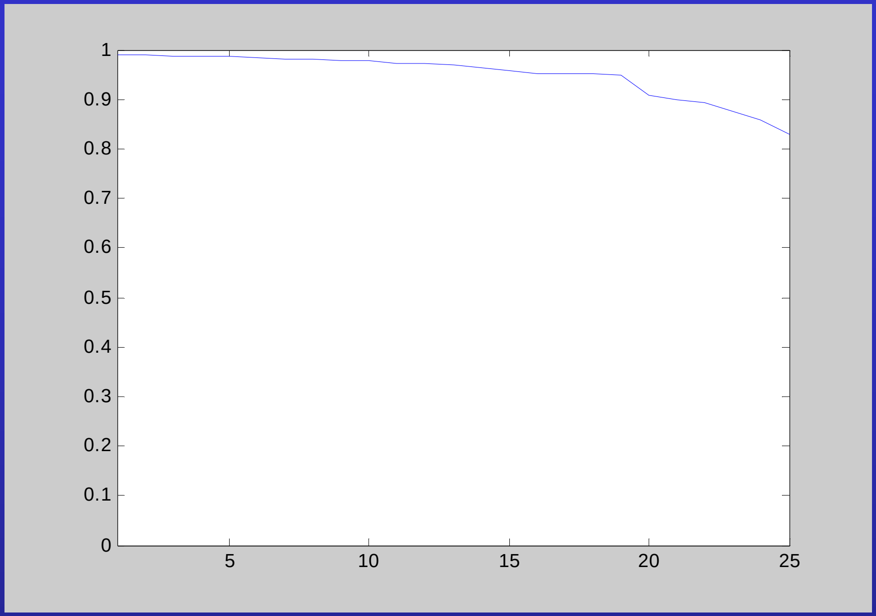


Red Plane

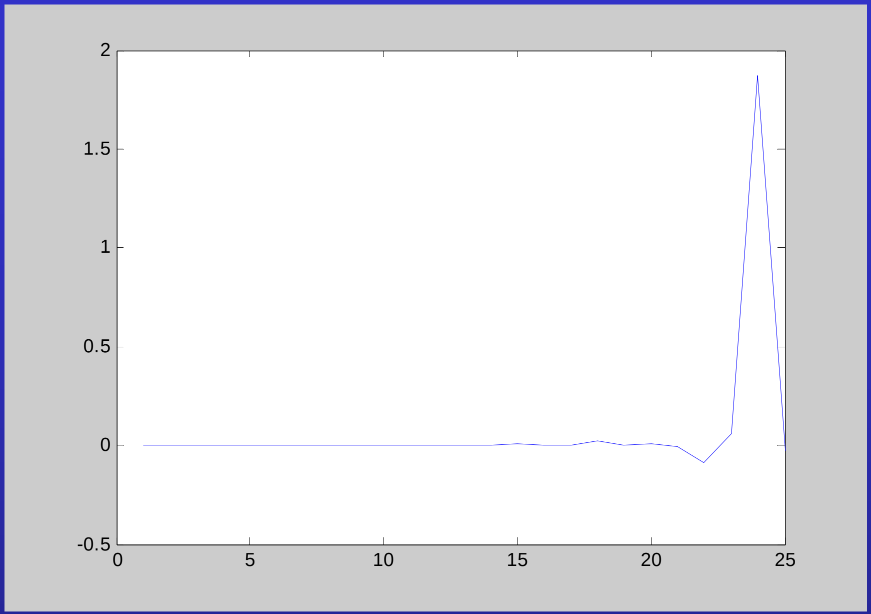


Green Plane

The 25 CC vectors



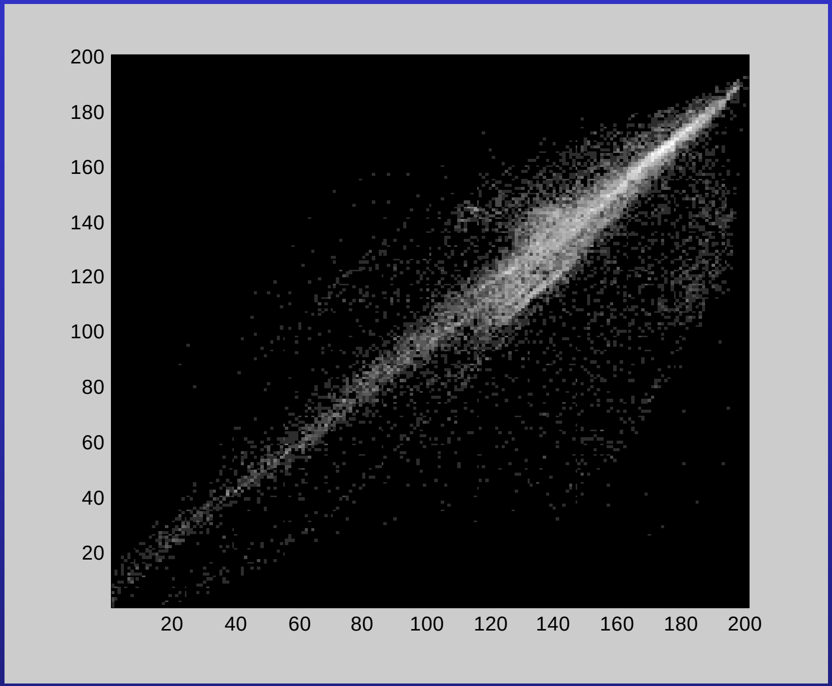
The 25 CC values for the CC vectors



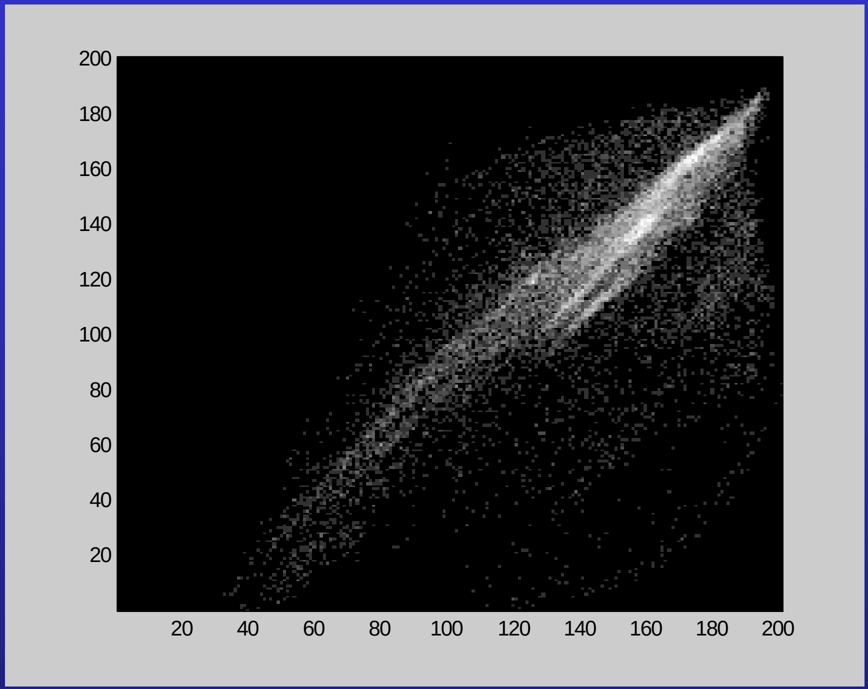
The 25 DC values for the CC vectors

- Observations:
  - The histogram shapes are highly stable over different images.
  - CC directions in 1x2 and 2x1 neighborhoods are the x-derivative and y-derivatives.
  - For kxk neighborhoods:
    - ➢ In all but one direction the DC value is zero.
    - ➢ In all DC=0 directions the CC values are almost identical.
    - ➢ Any linear combination of CC vectors with identical CC values, is a legitimate CC vector.

Comparison with PCA 1^{st} PC



Projected Red
v.s.
Projected Green



Projected Red
v.s.
Projected Blue

Comparison with PCA 1st PC

Subspace	Correlation	Entropy	Mutual Inf.	Cond. Ent.
Pure Spectral	0.91	8.60	1.75	6.84
PCA 2×1	0.94	8.19	1.78	6.40
PCA 1×2	0.94	8.24	1.76	6.48
PCA 3×3	0.94	8.13	1.86	6.26
CCA 2×1	0.99	5.03	1.65	3.38
CCA 1×2	0.98	5.04	1.50	3.54
CCA 2×2	0.99	4.64	1.72	2.92
CCA 3×3	0.99	4.53	1.68	2.84

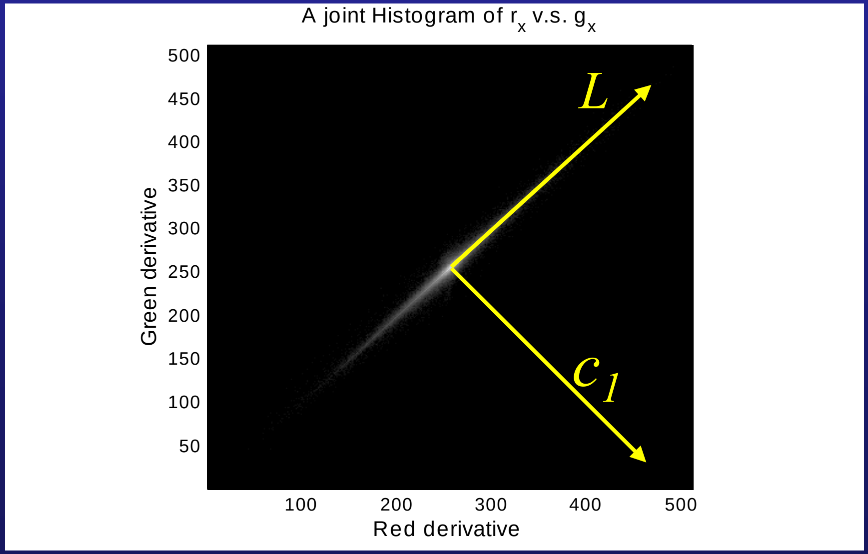
Table 1: Statistical values for various projected subspaces. All values were calculated for the Red and Green bands, and were averaged over 20 different natural images. The statistical values are (left to right): a. The correlation between Red and Green values: Corr(R,G). b. The differential entropy $H(R,G)$. c. The mutual information $I(R,G) = H(R,G) - H(R) - H(G)$. d. Two sided conditional entropy $H(R|G) + H(G|R) = H(R,G) - I(R,G)$.

Color Image Representation

- From now on we consider *x* and *y* derivatives as the CC directions (high pass filters).
- We define a new color basis (l, c_1, c_2):

$$\begin{pmatrix} l \\ c_1 \\ c_2 \end{pmatrix} = T \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad where \quad T = n \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & -2 \end{pmatrix}$$

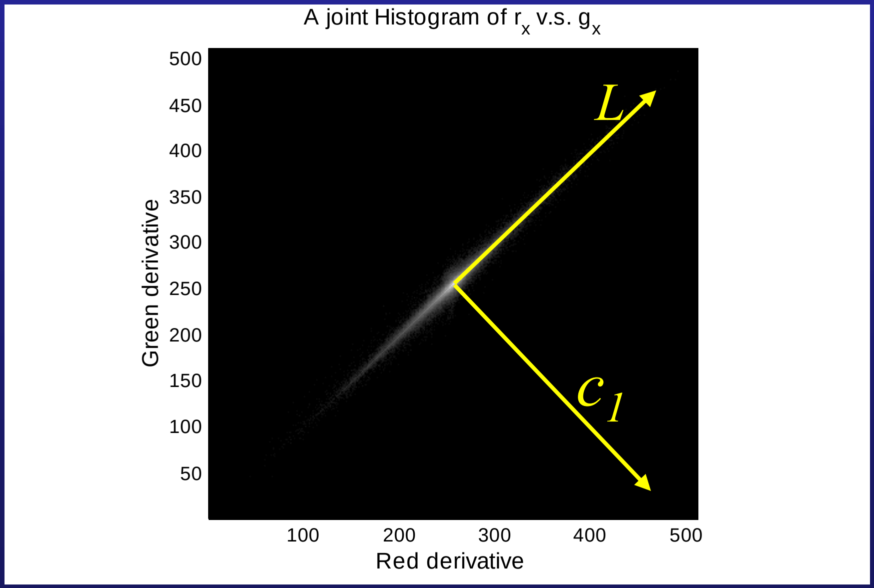
l – luminance
C_1- red/green
C_2 – blue/yellow



- Since spatial derivative is commutative we have that (l_x, c_{1x}, c_{2x}) is a rotated version of (R_x, G_x, B_x):

$$T\begin{pmatrix} R_x \\ G_x \\ B_x \end{pmatrix} = T\frac{\partial}{\partial x}\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \frac{\partial}{\partial x}T\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \frac{\partial}{\partial x}\begin{pmatrix} l \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} l_x \\ c_{1x} \\ c_{2x} \end{pmatrix}$$

- In the new coordinate system:
  - It is improbable to have high derivative values in the chrominance components.
  - It is probable to have high derivative values in the luminance component.



High derivatives

Low derivatives

Low derivatives

Claim: The HVS’ high spatial sensitivity in the Luminance domain and low spatial sensitivity in the Chrominance domains *is an outcome of the statistical properties of color images!*

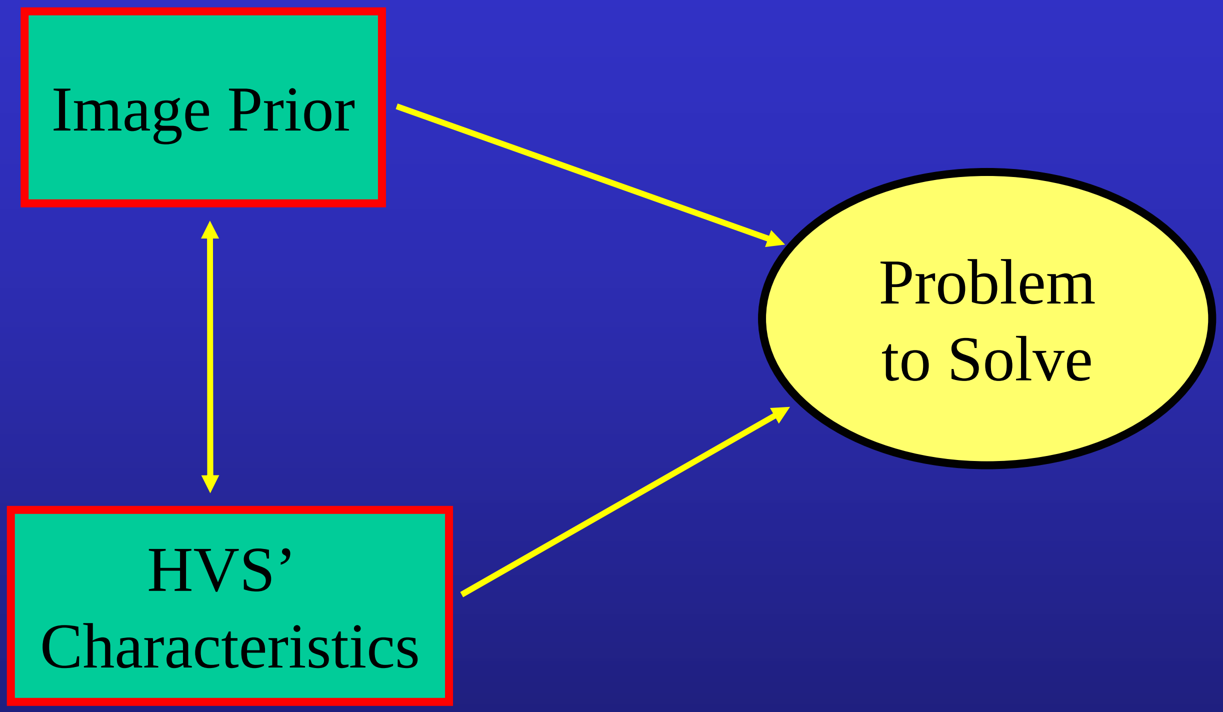


Image Prior
HVS' Characteristics
Problem to Solve

Parametric Prior for Color Images

- Useful Observation: The marginals in the $(l_x, C_{1,x}, C_{2,x})$ basis are statistically independent.
- Therefore it is possible to represent the p.d.f. as a product of marginals:

$$P(l_x, c_{1,x}, c_{2,x}) = P(l_x)P(c_{1,x})P(c_{2,x})$$

- Similarly for $(l_y, C_{1,y}, C_{2,y})$:

$$P(l_y, c_{1,y}, c_{2,y}) = P(l_y)P(c_{1,y})P(c_{2,y})$$

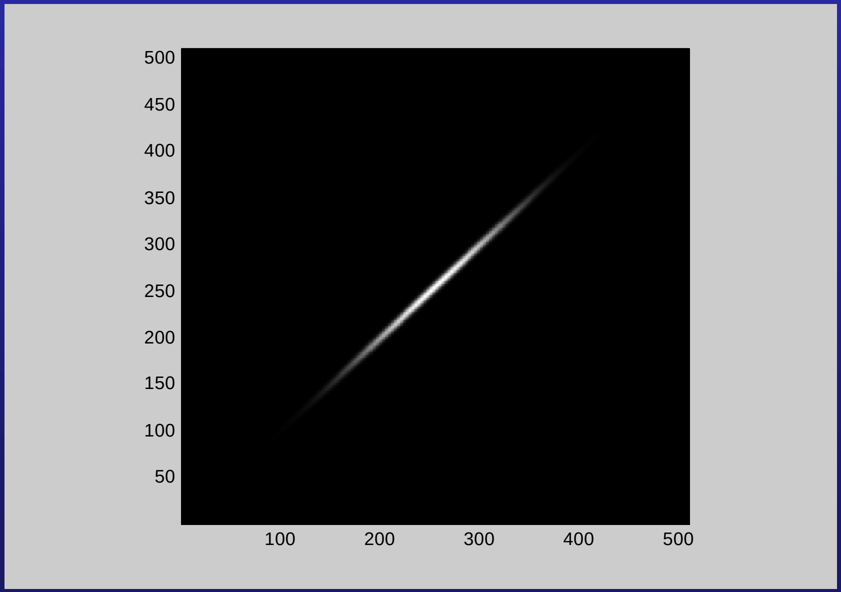
- Assuming x-derivatives and y-derivatives are stat. independent, we estimate the prior of color images in the CC subspaces:

$$P(R, G, B) \approx P(l_x, c_{1,x}, c_{2,x})P(l_y, c_{1,y}, c_{2,y})$$

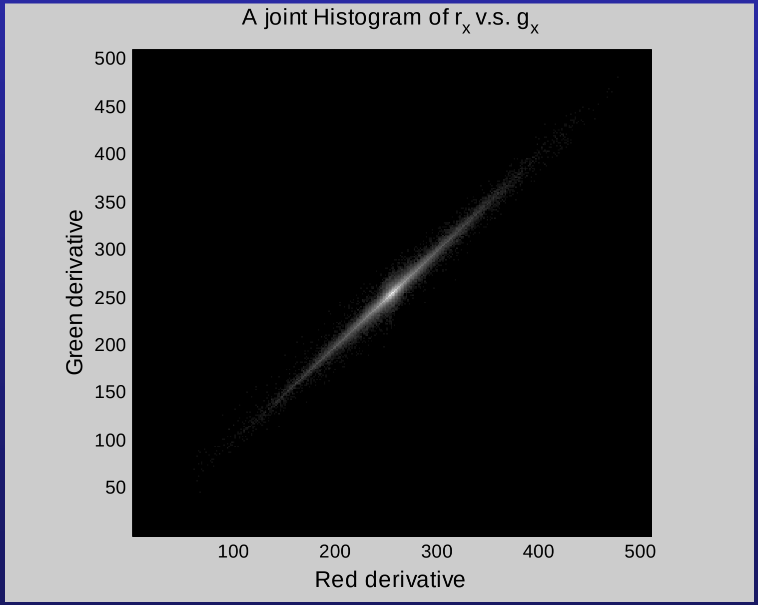
- First Approximation – A Joint Gaussian

$$P\left(c_{1_x}(x,y)\right) \propto \exp\left\{-\alpha^2 c_{1,x}^2(x,y)\right\}$$
$$P\left(c_{2_x}(x,y)\right) \propto \exp\left\{-\alpha^2 c_{2,x}^2(x,y)\right\}$$
$$P\left(l_x(x,y)\right) \propto \exp\left\{-\beta^2 l_x^2(x,y)\right\}$$

- The p.d.f. For $\alpha=1/5$ and $\beta=1/120$



Parametric Estimation



Actual Histogram

- Let $\mathbf{h}(x,y)$ be the image in the new basis:

$$\mathbf{h}(x,y) = T\,\mathbf{f}(x,y)$$

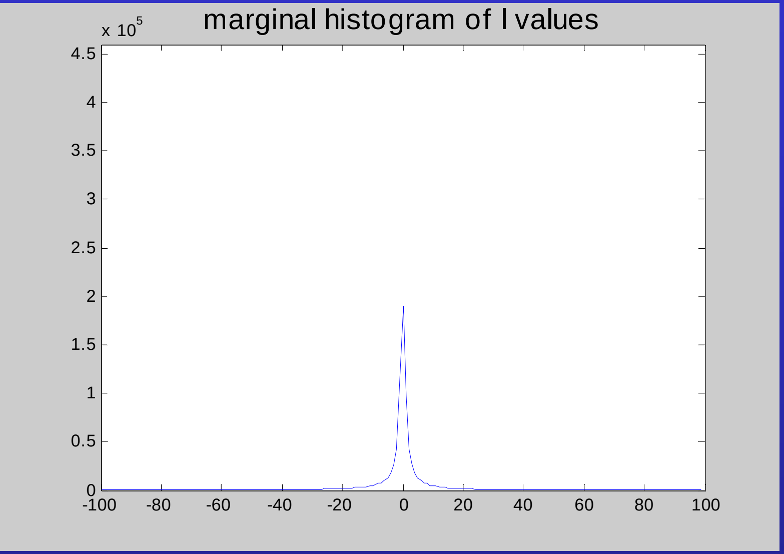
- Using the joint Gaussian model:

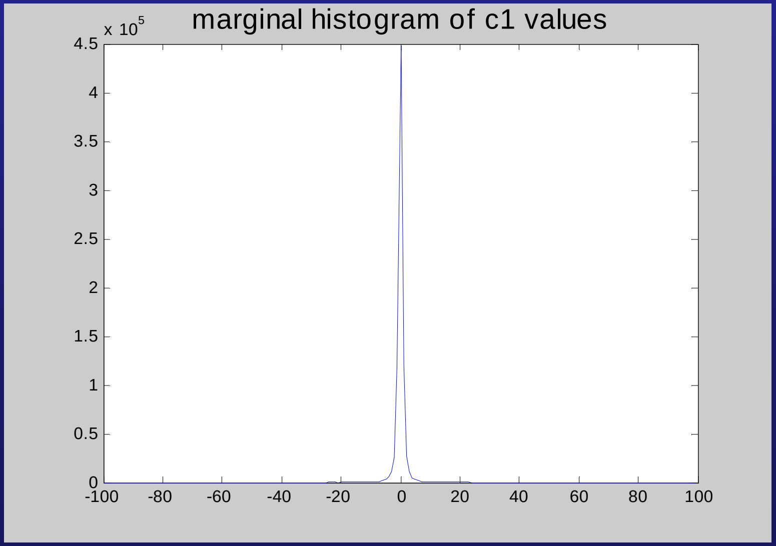
$$P_H(\mathbf{h}(x,y)) = \frac{1}{K}\exp\left\{-\alpha^2\left\|\nabla c_1(x,y)\right\|^2 - \alpha^2\left\|\nabla c_2(x,y)\right\|^2 - \beta^2\left\|\nabla l(x,y)\right\|^2\right\}$$

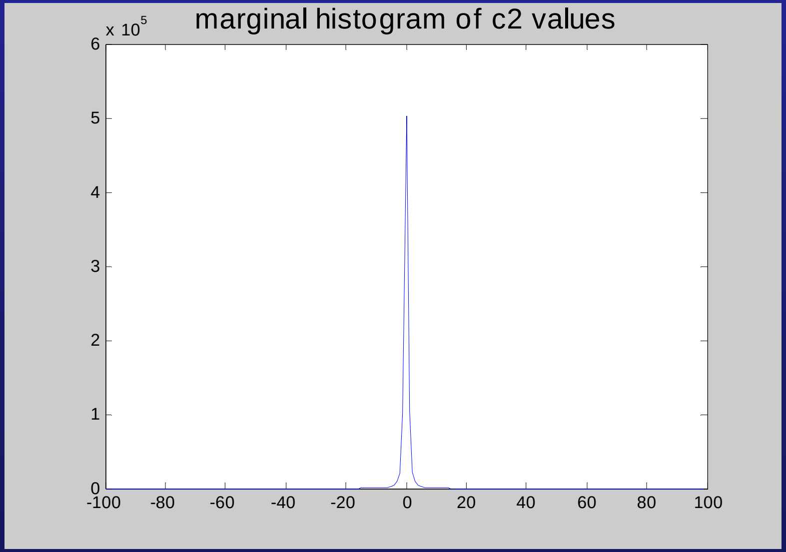
where

$$\nabla = \begin{pmatrix} \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} \end{pmatrix}^T$$

- Note: The Joint Gaussian model is convenient when applying it to inverse problems, however the marginals are Super-Gaussians:







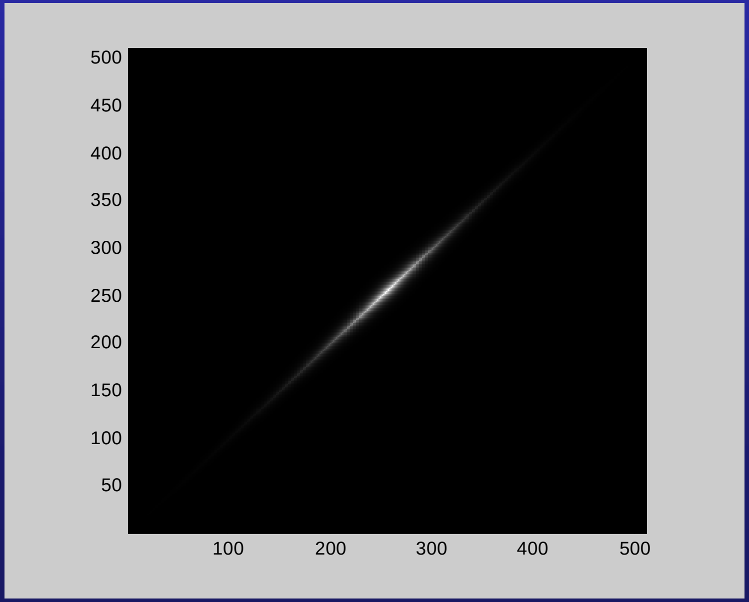
- Better Approximation – A Joint Laplacian

$$P\left(c_{1_x}(x,y)\right) \propto \exp\left\{-\alpha\left|c_{1,x}(x,y)\right|\right\}$$

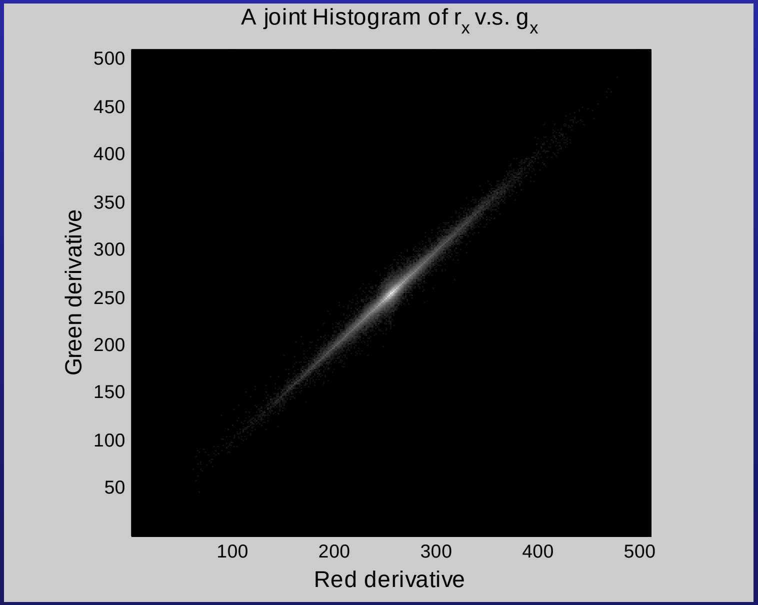
$$P\left(c_{2_x}(x,y)\right) \propto \exp\left\{-\alpha\left|c_{2,x}(x,y)\right|\right\}$$

$$P\left(l_x(x,y)\right) \propto \exp\left\{-\beta\left|l_x(x,y)\right|\right\}$$

- The p.d.f. For α=1/5 and β=1/80



Parametric Estimation



Actual Histogram

- Let $\mathbf{h}(x,y)$ be the image in the new basis:

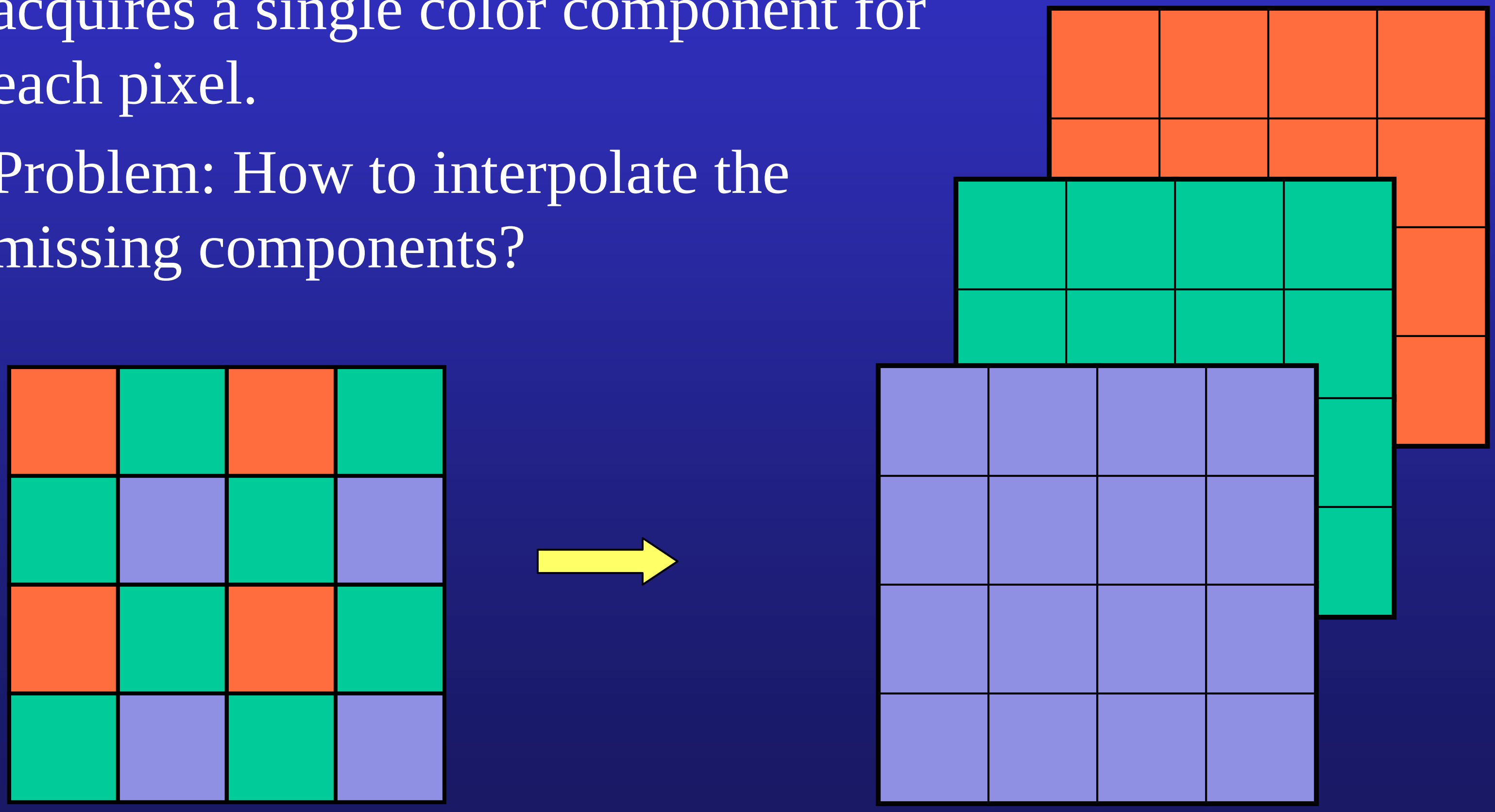
$$\mathbf{h}(x,y) = T\,\mathbf{f}(x,y)$$

- Using the joint Laplacian model:

$$\mathrm{P_H}\left(\mathrm{h}(\mathrm{x},\mathrm{y})\right) = \frac{1}{K} exp\left\{-\alpha\left\|\nabla c_1(x,y)\right\| - \alpha\left\|\nabla c_2(x,y)\right\| - \beta\left\|\nabla l(x,y)\right\|\right\}$$

Inverse Problem: Image Demosaicing

- The CCD sensor in a digital camera acquires a single color component for each pixel.
- Problem: How to interpolate the missing components?



Inverse Problem: Image Demosaicing

- Degradation Model:

$$m(x,y) = \mathbf{s}(x,y) \cdot \mathbf{f}(x,y)$$

  - ***f****(x,y)* the original image in the RGB basis
  - *m*(x,y) the mosaic image
  - **s**(x,y) a sampling mask

$$\mathbf{s}(x,y) = \begin{pmatrix} s^R(x,y) \\ s^G(x,y) \\ s^B(x,y) \end{pmatrix} \quad \textit{where} \qquad s^w(x,y) = \begin{cases} 1 & \textit{if color } \mathbf{w} \textit{ is sampled} \\ 0 & \textit{otherwise} \end{cases}$$

- Using MAP estimator, the optimal solution satisfies:

$$\hat{\mathbf{h}}(x,y) = \underbrace{\arg\max_h P_H(\mathbf{h})}_{\text{prior term}} \quad s.t. \quad \underbrace{m(x,y) = \mathbf{s}\cdot T^{-1}\mathbf{h}(x,y)}_{\text{data term}}$$

- Joint Gaussian case: Minimization is applied using POCS:
  - Minimizing the prior term:

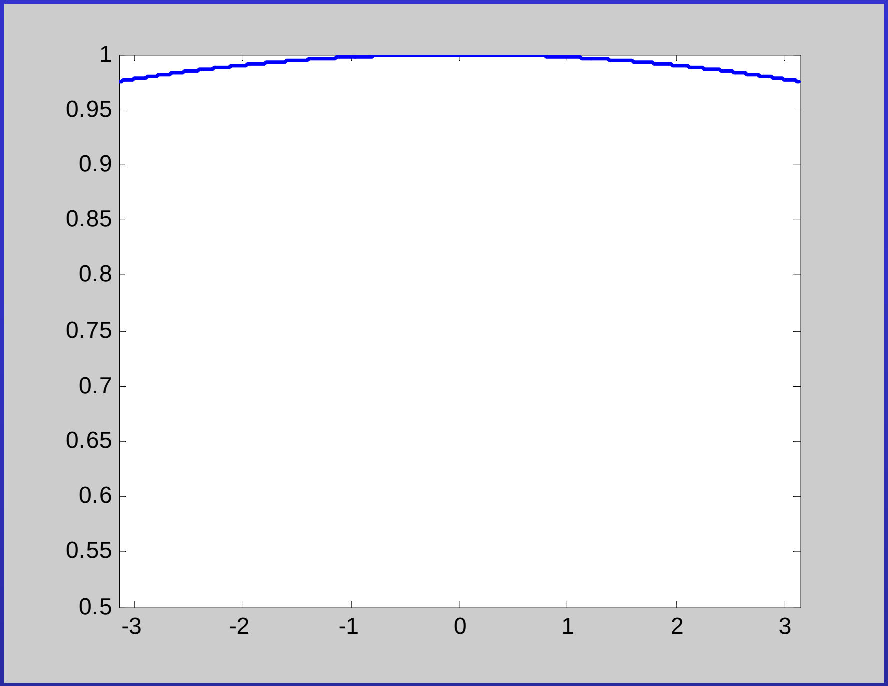
$$\min \sum_{x,y} \alpha \left\|\nabla c_1(x,y)\right\|^2 + \alpha \left\|\nabla c_2(x,y)\right\|^2 + \beta \left\|\nabla l(x,y)\right\|^2$$

$$l^{t+1}(x,y) = l^t(x,y) - \beta^2\nabla^2 l^t(x,y) = l^t(x,y) * \left(\delta - \beta^2\nabla^2\right)$$

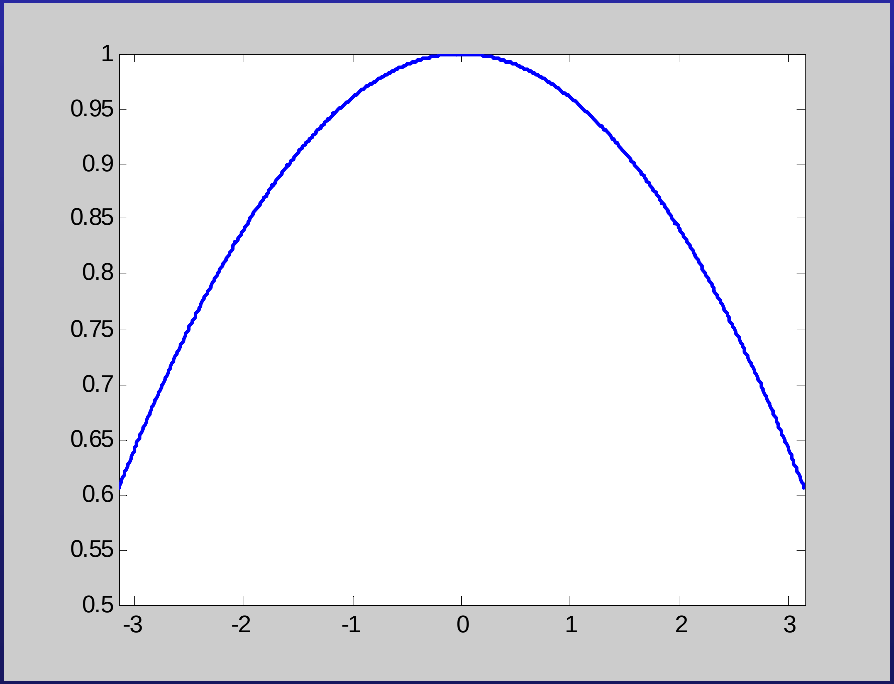
$$c_i^{t+1}(x,y) = c_i^t(x,y) - \alpha^2\nabla^2 c_i^t(x,y) = c_i^t(x,y) * \left(\delta - \alpha^2\nabla^2\right)$$

  - Projecting temporary results onto the constraint:

$$m(x,y) = \mathbf{s}\cdot T^{-1}\mathbf{h}(x,y)$$

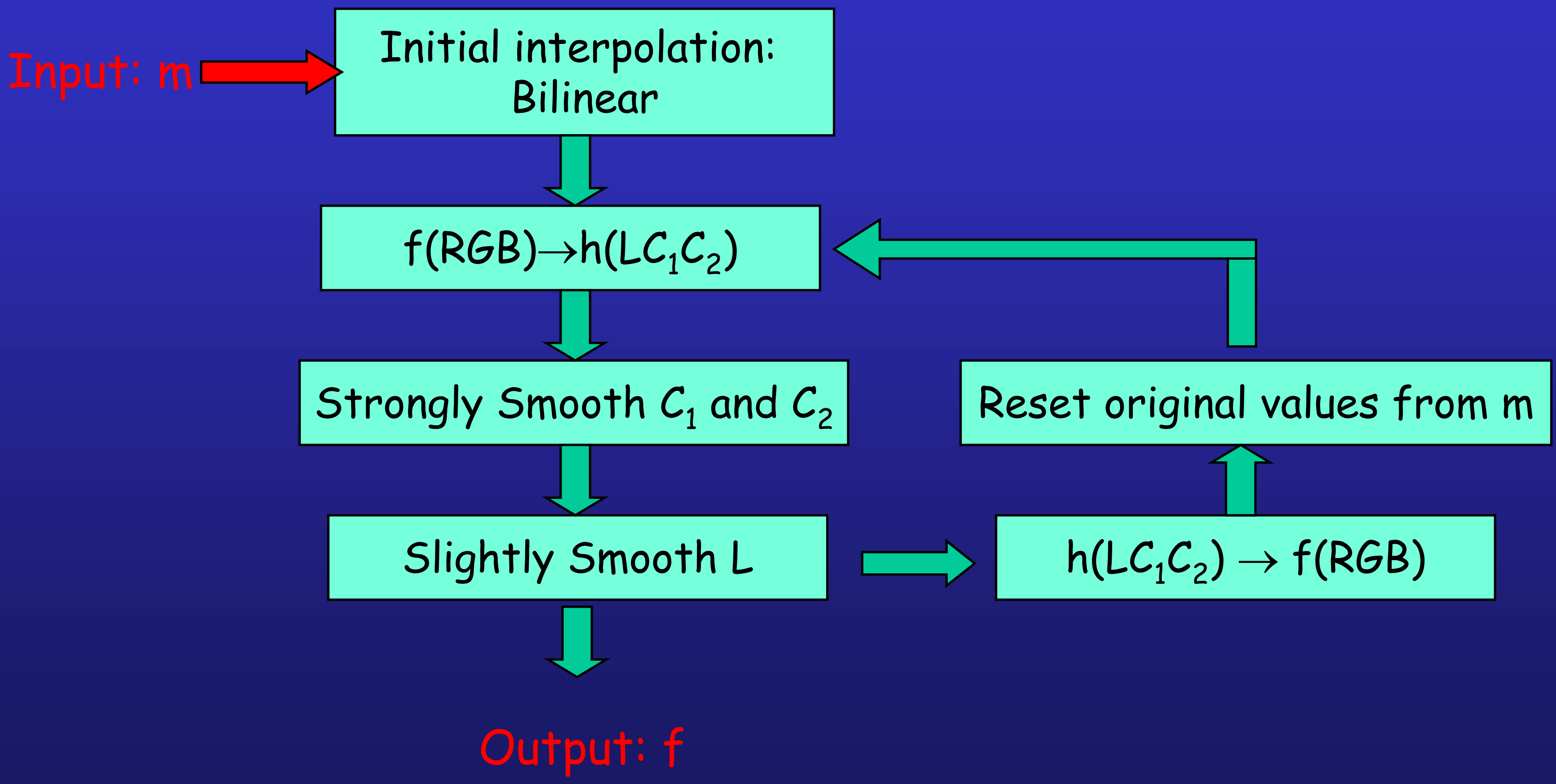


Freq. Profile of the kernel applied to the Luminance channels.



Freq. Profile of the kernel applied to the Chrominance channels.

The Demosaicing Algorithm (Gaussian Model)



Remarks

- Since each iteration applies a linear operation, the entire algorithm can be implemented using a single linear operation.
- A similar algorithm can be produced using only the characteristics of the HVS where minimizing $P_H(\nabla \mathbf{h}(x,y))$ can be interpreted as a perceptual penalty.
- Modeling $P_H(h)$ using a joint Laplacian p.d.f. yields an adaptive filtering which preserves edges (TV norm for Laplacian distribution).

$$\min \sum_{x,y} \alpha\left|\nabla c_1(x,y)\right| + \alpha\left|\nabla c_2(x,y)\right| + \beta\left|\nabla l(x,y)\right|$$

$$\Rightarrow \quad I^{t+1}(x,y) = I^t(x,y) - \beta^2 \frac{\nabla^2 I^t(x,y)}{\left|\nabla I^t(x,y)\right|} \quad ; \quad c_i^{t+1}(x,y) = c_i^t(x,y) - \alpha^2 \frac{\nabla^2 c_i^t(x,y)}{\left|\nabla I^t(x,y)\right|}$$

Initial Solution

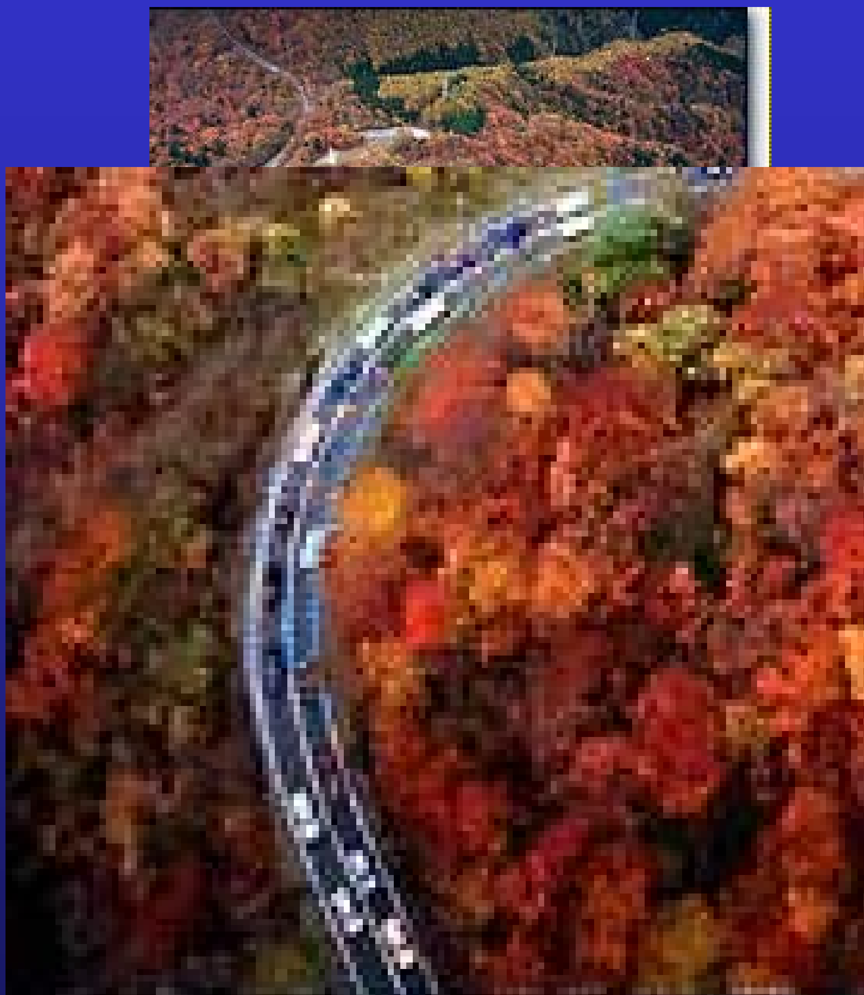
Initial Solution

Final Solution

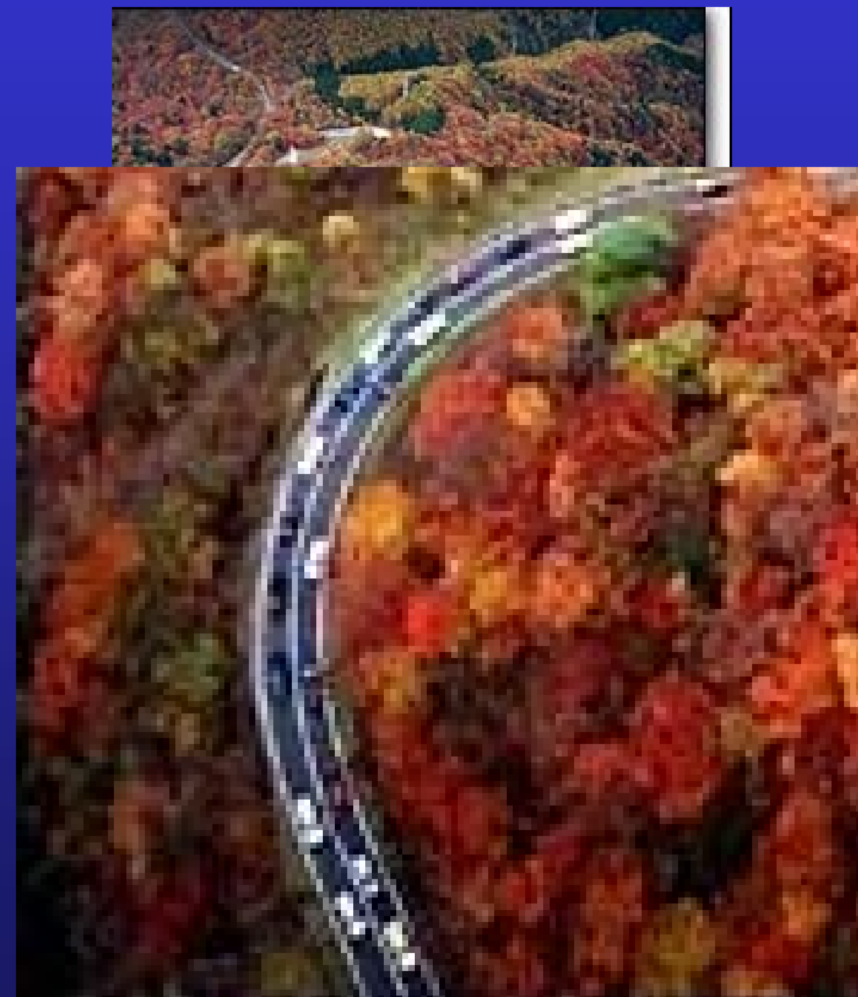
Initial Solution

Final Solution

Demosaicing Results using adaptive filtering



Optimal Linear Demosaicing

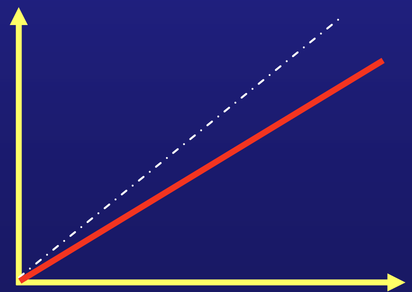


Adaptive CC Demosaicing

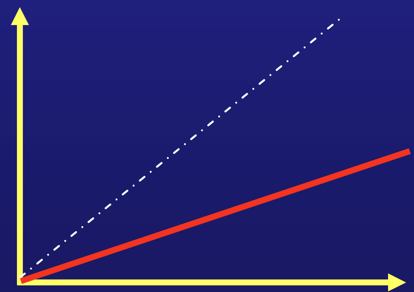
Conclusions

- Modeling the prior of natural color images is important.
- Modeling the prior in the Canonical Correlation directions is useful in some applications (e.g. demosaicing).
- The statistical properties of color images in the CC directions conforms with the characteristics of the HVS.

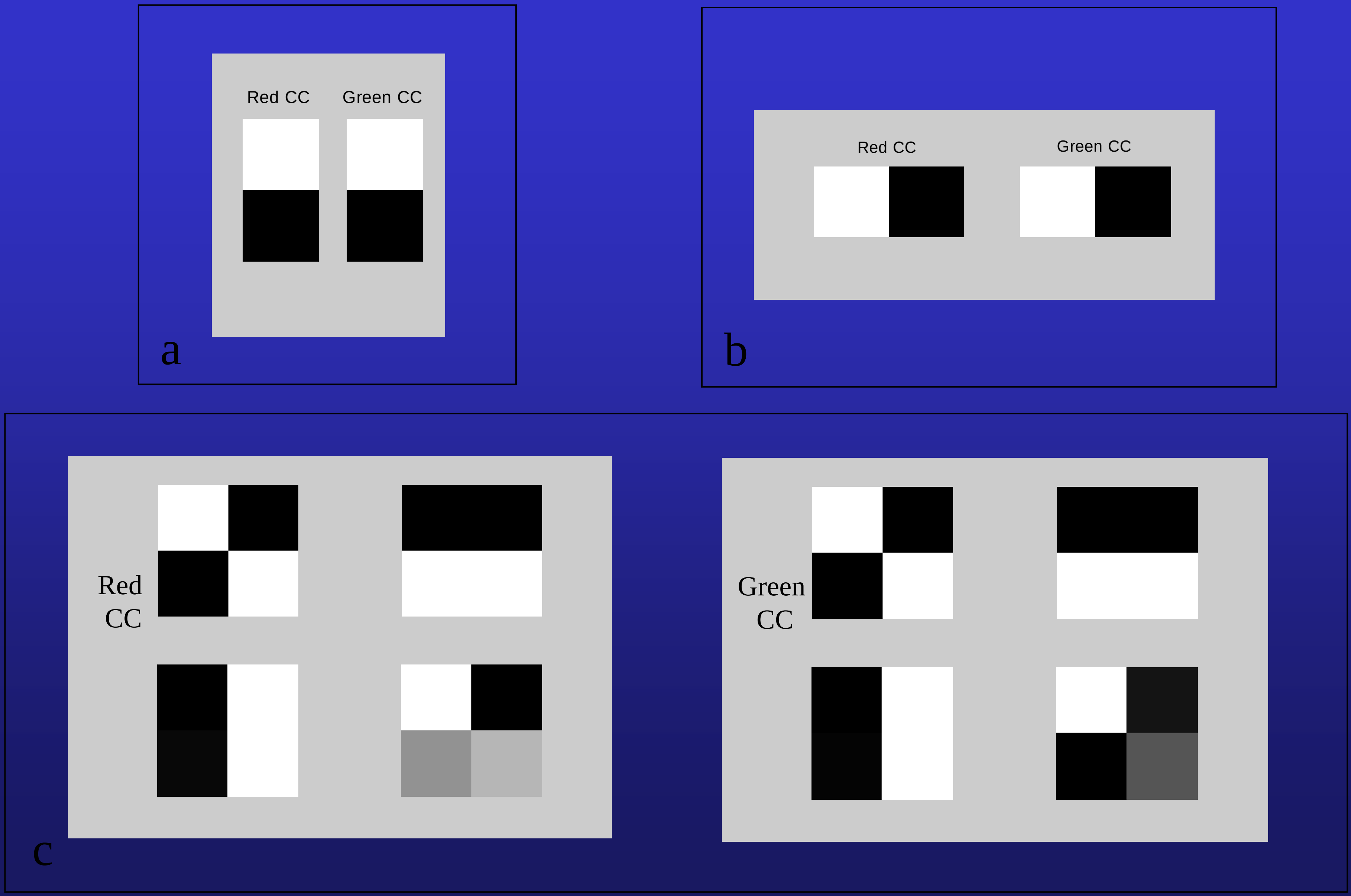
- Future Work:
  - Noise removal using Soft / Hard thersholding in the CC basis

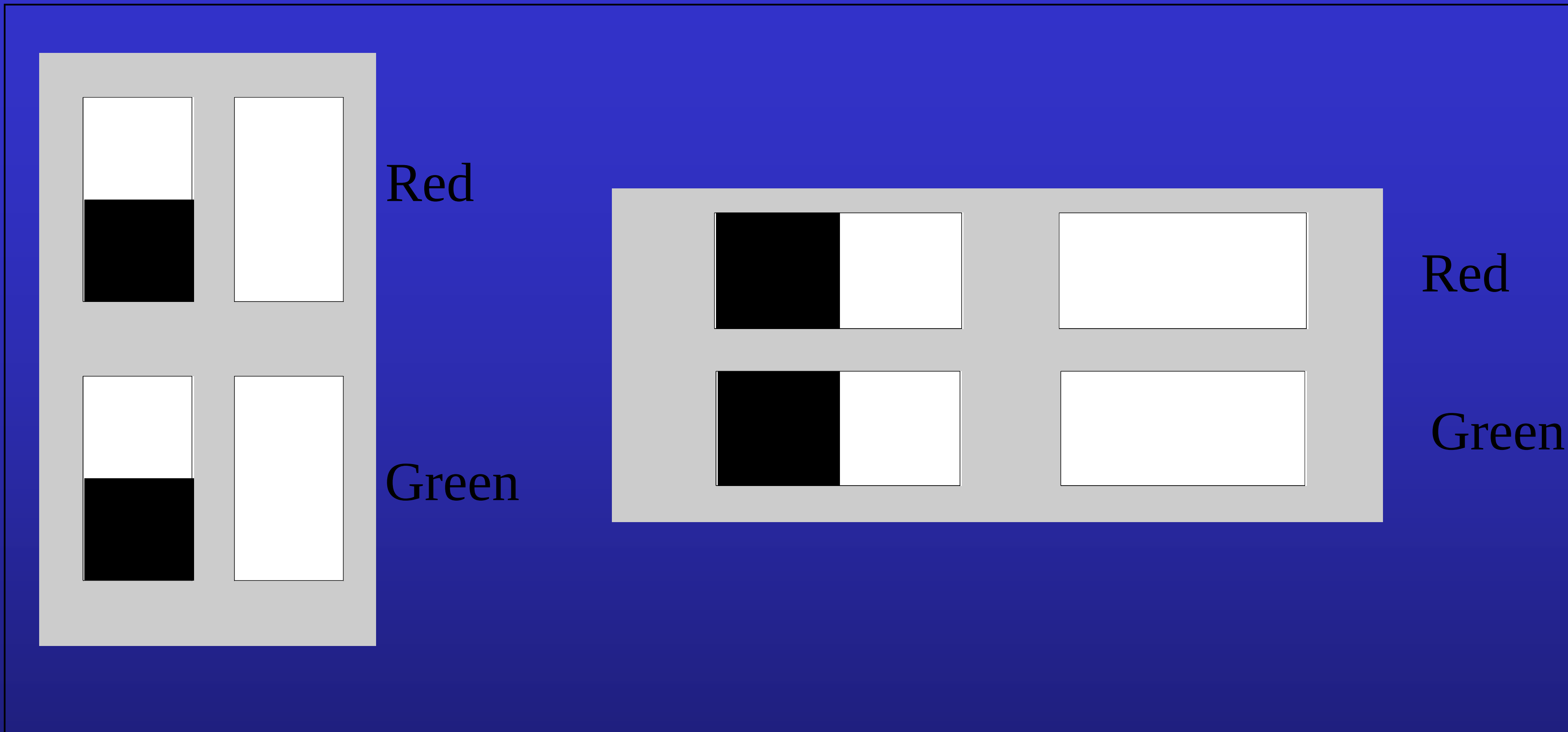


CC luminance



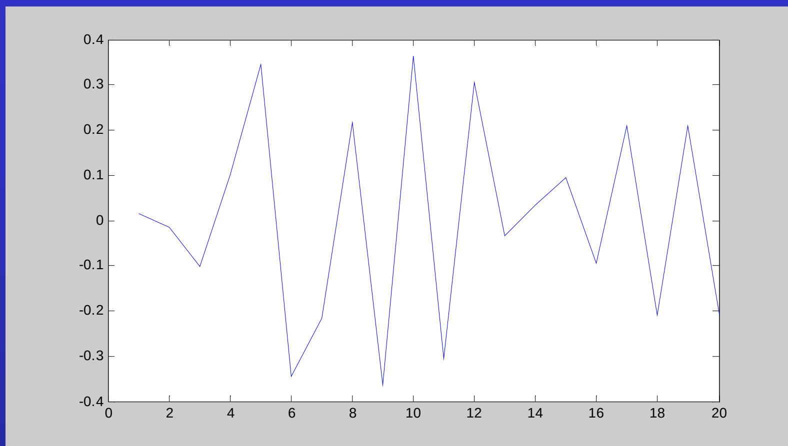
CC chrominance


Red CC
Green CC
a
Red CC
Green CC
b
Red CC
Green CC
c

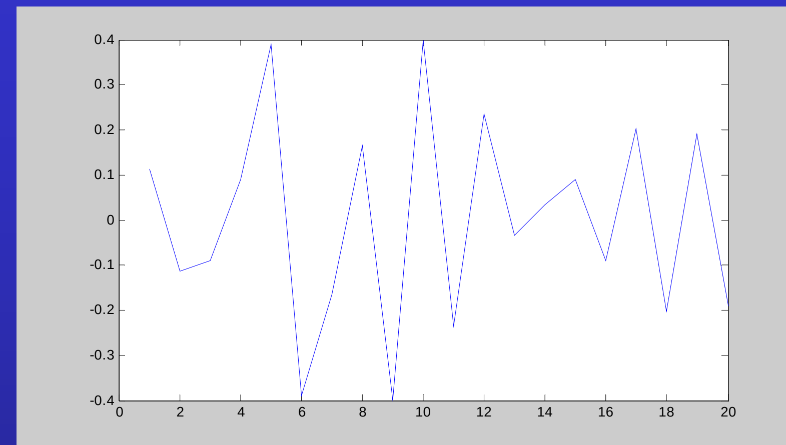

Red
Green
Red
Green

Example: $\mathbf{x}$ and $\mathbf{y}$ are 20D random vectors where:

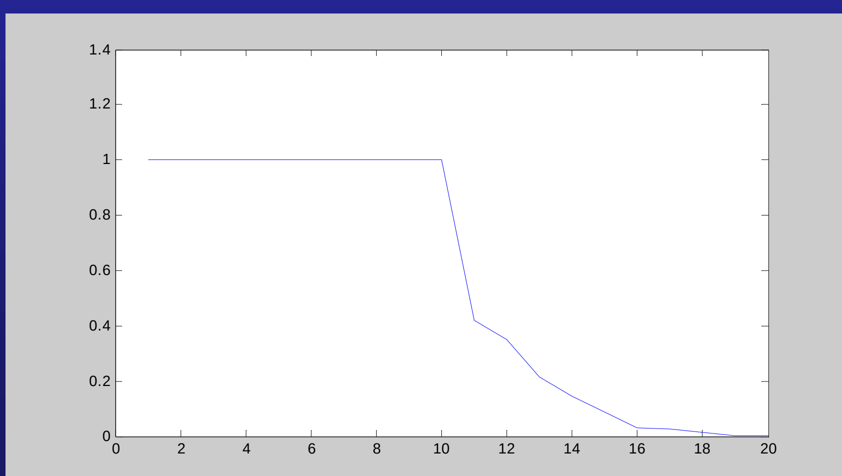
$$\mathbf{x}(2i)-\mathbf{x}(2i-1)=\mathbf{y}(2i)-\mathbf{y}(2i-1)$$



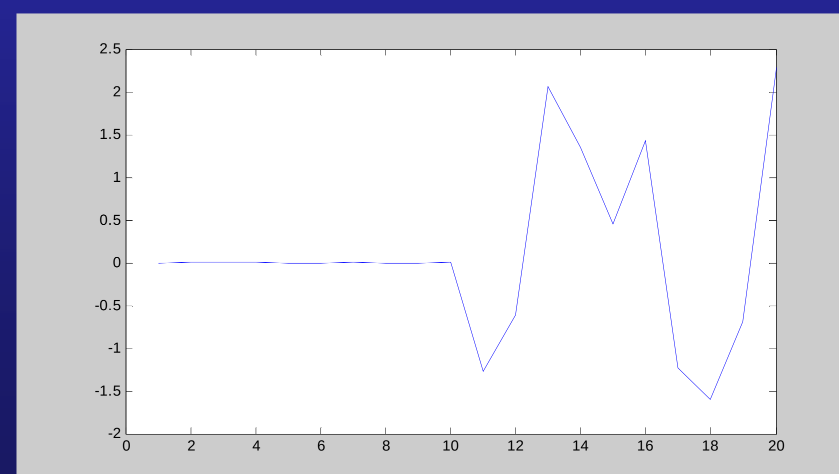
1st CC vector of x



1st CC vector of y



The CC values



The DC of the CC vectors