CS754 Assignment 5

Hitesh Kumar Punna, Sudhansh Peddabomma

190050093, 190050118

Question 1.

- Equation 6 is

$$J_2(I_1) = \sum_{i,k} \rho(f_{i,k} \cdot I_1) + \rho(f_{i,k} \cdot (I - I_1)) + \lambda \sum_{i \in S_1,k} \rho(f_{i,k} \cdot I - f_{i,k} \cdot I_1) + \lambda \sum_{i \in S_2,k} \rho(f_{i,k} \cdot I_1)$$

Equation 7 corresponds to the same cost minimization function and is written as,

$$J_3(v) = \sum_j (A_{j\rightarrow} v - b_j)$$

Here, v is the vectorized form of I_1 and $A_{j\rightarrow}$ is the matrix representing the derivative filter. b_j is either vectorized $f \cdot I$ or 0 depending upon j.
We can obtain such an $A_{j\rightarrow}$ because convolution is a linear operation and $f_{i,k} \cdot I_1$ is the result of application of derivative filter on I_1. Let M be the matrix representation of derivative filter f for the *whole image*.

Each corresponding column of M is set to 0 when $i \in S_1$ and $i \in S_2$ are being considered. The following table summarises the meaning of $A_{j\rightarrow}$ and b_j for each of the terms in Equation 6 below.

No	Term	$A_{j\rightarrow}$	b_j
1	$\sum_{i,k} \rho(f_{i,k} \cdot I_1)$	M	0
2	$\sum_{i,k} \rho(f_{i,k} \cdot (I - I_1))$	-M	M$\times I$
3	$\lambda \sum_{i \in S_1,k} \rho(f_{i,k} \cdot I - f_{i,k} \cdot I_1)$	-M_{S_1}	$M_{S_1} \times I$
4	$\lambda \sum_{i \in S_2,k} \rho(f_{i,k} \cdot I_1)$	M_{S_2}	0

where M_{S_i} represents M with columns corresponding to S_i set to 0.

- The prior over images is given by

$$Pr(I) \approx \prod_{i,k} Pr(f_{i,k} \cdot I)$$

where $f \cdot I$ denotes the inner product between linear filter f and an image I, and $f_{i,k}$ is the kth derivative filtered centered on pixel i. The derivative filters set we use includes two orientations (horizontal and vertical) and two degrees (i.e. first derivative filters as well as second derivative).

The likelihood of the filter is given by the Laplacian mixture model. That is,

$$\log Pr(f_{i,k} \cdot I) \approx -rho(F_{i,k} \cdot I)$$

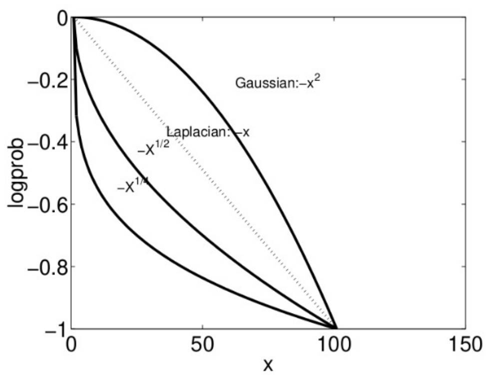
$$\rho(x) = log(\frac{\pi_1}{2s_1} e^{-|x|/s_1} + \frac{\pi_2}{2s_2} e^{-|x|/s_2})$$

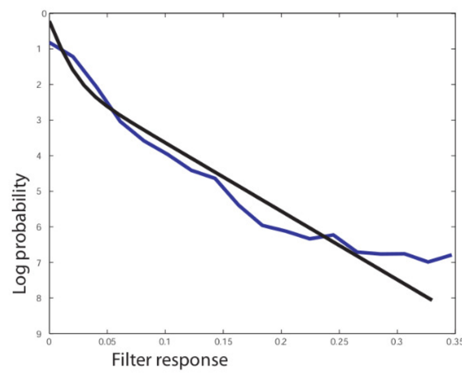
Therefore, the terms $\sum_{i,k} \rho(f_{i,k} \cdot I_1) + \rho(f_{i,k} \cdot (I - I_1))$ correspond to the prior and the terms $\lambda \sum_{i \in S_1,k} \rho(f_{i,k} \cdot I - f_{i,k} \cdot I_1) + \lambda \sum_{i \in S_2,k} \rho(f_{i,k} \cdot I_1)$ correspond to the likelihood of the derivative filter.

- We use Laplacian instead of Gaussian likelihood because the experimental distribution is closer to a Laplacian distribution. On plotting the coefficients of the matrix obtained by applying derivative filter on natural images, we get the following.

  Another reason for using the Laplacian prior is that users can make mistakes in S_1 and S_2, and absolute error is more robust than squared error in case of outliers.

  We can see that the values of the image is closer to a Laplacian (the straight line). Also, The Gaussian distribution is not sparse (it is always above the straight line) and distributions for which $\alpha < 1$ are sparse. Hence, we use a Laplacian mixture model.





Question 2.

We have

$$y = \phi x + \eta$$

It is given that x is drawn from a Gaussian distribution with mean 0 and covariance Σ_x with size $n \times n$. From Conditional probability, we have

$$P(x|y,\phi,C_x) = \frac{P(y|x,\phi,C_x)P(x)}{P(y)} \propto P(y|x,\phi,C_x)P(x)$$

We can work on $P(y|x,\phi,C_x)P(x)$ as the MAP estimate does not change due to a constant factor. Now, we have

$$P(x) = \frac{exp(-x^T C_x^{-1} x/2)}{(2\pi)^{n/2}|C_x|^{1/2}}$$

as x is sampled from a 0 mean multivariate Gaussian. Also, $P(y|x,\phi,C_x)$ follows the distribution of $\eta \sim \mathcal{N}(0, \sigma^2 I_{m\times m})$ which is given by

$$P(y|x,\phi,C_x) = \frac{exp(-\left\|y-\phi x\right\|_2^2/2\sigma^2)}{(2\pi\sigma^2)^{m/2}}$$

Now,

$$P(x|y,\phi,C_x) \propto \frac{exp(-x^T C_x^{-1} x/2)}{(2\pi)^{n/2}|C_x|^{1/2}} \frac{exp(-\left\|y-\phi x\right\|_2^2/2\sigma^2)}{(2\pi\sigma^2)^{m/2}} \propto exp(-x^T C_x^{-1} x/2) exp(-\left\|y-\phi x\right\|_2^2/2\sigma^2)$$

We can consider the Log-likelihood to get the MAP estimate

$$\log P(x|y,\phi,C_x) = -x^T C_x^{-1} x/2 + -\left\|y-\phi x\right\|_2^2/2\sigma^2 + \text{constants independent of } x$$

$$\hat{x}_{MAP} = arg\ min_x x^T C_x^{-1} x/2 + \left\|y-\phi x\right\|_2^2/2\sigma^2$$

$$\frac{\partial \log P(x|y,\phi,C_x)}{\partial x} = (y-\phi x)(-\phi/\sigma^2) + x^T C_x^{-1} = 0 \implies x^T C_x^{-1} = (y-\phi x)^T \phi/\sigma^2$$

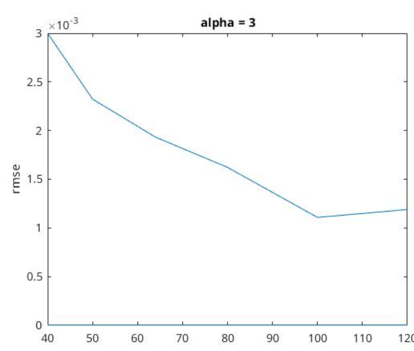
Transpose on both sides,

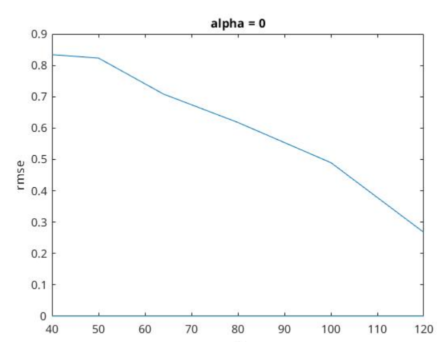
$$\sigma^2 C_x^{-T} x = \phi^T(y-\phi x) \implies (\sigma^2 C_x^{-T} + \phi^T\phi)x = \phi^T y$$

Therefore, the MAP estimate is given by,

$$\hat{x}_{MAP} = (\sigma^2 C_x^{-T} + \phi^T\phi)^{-1}\phi^T y$$

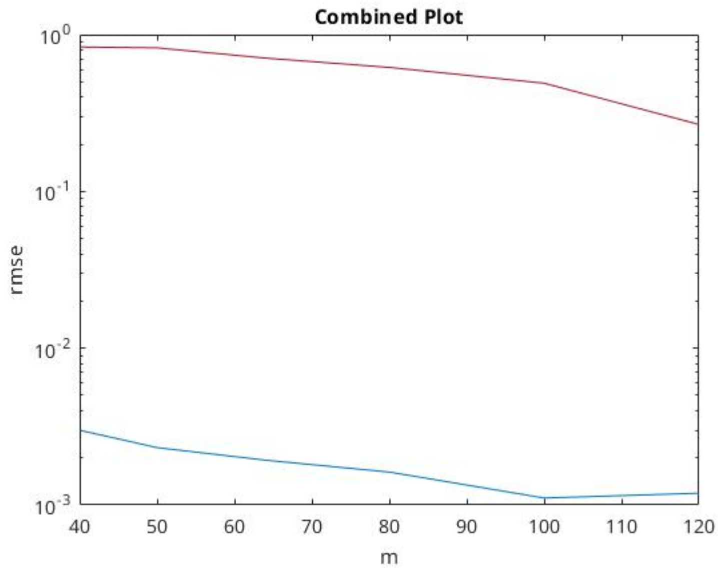
The simulation was performed and the following results were obtained. The code is present in `q2.m`.





Firstly, we notice that as m increases, RMSE decreases as expected. Now, when α is reduced, we notice that RMSE decreases. This is because of the fact that the singular values of C_x are of the form $c \cdot i^{-\alpha}$. When α is 0, all the coefficients are equal ($= 1$), and the signal is incompressible. Although, when $\alpha = 3$, some coefficients are much smaller than the others and the signal can be compressed. That is, the space can be represented by a much smaller set of vectors. Therefore, y can capture a lot more information when α is higher. *Note.* The definition of RMSE used here is $\sqrt{\texttt{mean}((\hat{x} - x)^2)}$

The combined plot is shown below. *Note.* y axis is in logarithmic scale.



Question 3.

(a) We are finding X^* in the following way

$$X^* = argmin_X \|X\|_*$$

such that $AX = b$
What this means is for all Y satisfying $AY = b$,

$$\|X^*\|_* \leq \|Y\|_*$$

As we know $AX_0 = b$, from the previous equation we can say that

$$\|X^*\|_* \leq \|X_0\|_*$$

(b) As we can see from Lemma 3.4 that there exist matrices R and R_c such that $X_0'R_c = 0$ and $X_0R_c' = 0$
Now using Lemma 2.3 which is, Let A and B be matrices of the same dimensions. If $AB' = 0$ and $A'B = 0$ then $\|A + B\|_* = \|A\|_* + \|B\|_*$
As $R = X^* - X_0$ and $R = R_0 + R_c$, the dimension of X_0 and R_c are same.
From this we can see that if $A = X_0, B = R_c \implies \|X_0 + R_c\|_* = \|X_0\|_* + \|R_c\|_*$
As we already know

$$\|X_0\|_* \geq \|X_0 + R\|_* \geq \|X_0 + R_c\|_* - \|R_0\|_*$$

From the previous equality we can write

$$\|X_0\|_* \geq \|X_0 + R\|_* \geq \|X_0 + R_c\|_* - \|R_0\|_* = \|X_0\|_* + \|R_c\|_* - \|R_0\|_*$$

(c) As we know by definition that

$$\sigma_k \leq \sigma_j \ \ \forall j \in I_i \ \ \forall k \in I_{i+1}$$

On applying summation on both sides, we will get

$$\Sigma_{j \in I_i} \sigma_k \leq \Sigma_{j \in I_i} \sigma_j \ \ \forall k \in I_{i+1}$$

As the size of $I_i = 3r$ we can say that

$$3r * \sigma_k \leq \Sigma_{j \in I_i} \sigma_j \ \ \forall k \in I_{i+1}$$

In this way we can prove the inequality

$$\sigma_k \leq \frac{1}{3r} \Sigma_{j \in I_i} \sigma_j \ \ \forall k \in I_{i+1}$$

(d) As we know that $diag(\sigma_{I_{i+1}})$ contains singular values for matrix R_{i+1}, we can write

$$\|R_{i+1}\|_F^2 = \Sigma_{j \in I_{i+1}} \sigma_j^2$$

and from definition of nuclear norm we can write

$$\|R_i\|_*^2 = (\Sigma_{j \in I_i} \sigma_j)^2$$

on squaring the equation from part c) we will get

$$\sigma_k^2 \leq \frac{1}{9r^2} (\Sigma_{j \in I_i} \sigma_j)^2 \ \ \forall k \in I_{i+1}$$

which can be also represented as

$$\sigma_k^2 \leq \frac{1}{9r^2} \|R_i\|_*^2 \ \ \forall k \in I_{i+1}$$

on applying summation over all $k \in I_{i+1}$ we get

$$\Sigma_{k \in I_{i+1}} \sigma_k^2 \leq \Sigma_{k \in I_{i+1}} \frac{1}{9r^2} \|R_i\|_*^2$$

Now this can be reduced to

$$\|R_{i+1}\|_F^2 \leq \frac{1}{9r^2} \|R_i\|_*^2 \Sigma_{k \in I_{i+1}} 1$$

$$\|R_{i+1}\|_F^2 \leq \frac{1}{9r^2} \|R_i\|_*^2 * 3r$$

from this we can say that

$$\|R_{i+1}\|_F^2 \leq \frac{1}{3r} \|R_i\|_*^2$$

(e) As we saw in the previous step

$$\|R_{i+1}\|_F^2 \leq \frac{1}{3r} \|R_i\|_*^2$$

we see that it satisfies $\forall i \in \{1, 2, 3, ...\}$
Hence, we will apply summation over $i \in \{1, 2, 3, ...\}$ on both sides of the equation.From that we will get

$$\Sigma_{i \geq 1} \|R_{i+1}\|_F^2 \leq \Sigma_{i \geq 1} \frac{1}{3r} \|R_i\|_*^2$$

Now we on substituting $j = i + 1$ on Left term we will get

$$\Sigma_{i \geq 1} \|R_{i+1}\|_F^2 = \Sigma_{j \geq 2} \|R_j\|_F^2$$

From this we can show that

$$\Sigma_{j \geq 2} \|R_j\|_F^2 \leq \Sigma_{i \geq 1} \frac{1}{3r} \|R_i\|_*^2$$

This equation can be rewritten as

$$\Sigma_{j \geq 2} \|R_j\|_F^2 \leq \Sigma_{j \geq 1} \frac{1}{3r} \|R_j\|_*^2$$

(f) As we know

$$\|R_c\|_* = \Sigma_{i \geq 1} \sigma_i$$

and we know that set

$$\{i | i \geq 1\} = \bigcup_{i \geq 1} I_i$$

we can write

$$\|R_c\|_* = \Sigma_{i \geq 1} \sigma_i = \Sigma_{i \geq 1} \Sigma_{j \in I_i} \sigma_j$$

As we know $\|R_i\|_* = \Sigma_{j \in I_i} \sigma_j$, the above equation can be simplified to

$$\|R_c\|_* = \Sigma_{i \geq 1} \sigma_i = \Sigma_{i \geq 1} \|R_i\|_*$$

From part (b) of this question we saw that

$$\|X_0\|_* \geq \|X_0\|_* + \|R_c\|_* - \|R_0\|_*$$

from this we can say that

$$\|R_0\|_* \geq \|R_c\|_*$$

From the above two derived equations we can say that

$$\Sigma_{i \geq 1} \|R_i\|_* \leq \|R_0\|_*$$

This justifies the equation

$$\frac{1}{3r} \Sigma_{i \geq 1} \|R_i\|_* \leq \frac{1}{3r} \|R_0\|_*$$

(g) Let the rank of Matrix $R_0 = k$
As we applyin Lemma(3.4) to the matrices X_0 and R,to show that there exist matrices R_0 and R_c such that $R = R_0 + R_c$, Here matrix R_0 satisfies the condition $rank(R_0) \leq 2rank(X_0)$
As rank of matrix $X_0 = r \implies k \leq 2r$
Now let $R_0 = U_0 diag(S) V_0'$ upon singular value decomposition
And as we know

$$\|R_0\|_F^2 = \Sigma_{j\geq 1} S_j^2$$

where each S_j is singular value of matrix R_0
Similarly

$$\|R_0\|_* = \Sigma_{j\geq 1} S_j$$

As rank of a matrix = number of non-zero singular values of the matrix
As rank of $R_0 = k \implies$ Number of non-zero $S_j = k$
Let $i \in \{a_1, a_2, ..a_k\}$ be set of indices for which $S_i \neq 0$

$$\|R_0\|_F^2 = \Sigma_{i=1}^k S_{a_i}^2$$

$$\|R_0\|_* = \Sigma_{i=1}^k S_{a_i}$$

Now by cauchy schwarz inequality we can say that

$$(\Sigma_{i=1}^k S_{a_i})^2 \leq (\Sigma_{i=1}^k S_{a_i}^2) * k$$

This implies that

$$(\|R_0\|_*)^2 \leq \|R_0\|_F^2 * k$$

$$(\|R_0\|_*) \leq \|R_0\|_F^2 * \sqrt{k}$$

As we know $k \leq 2r$, we reduce the above equation to

$$\|R_0\|_* \leq \|R_0\|_F * \sqrt{2r}$$

This Prove the inequality

$$\frac{1}{3r}\|R_0\|_* \leq \frac{1}{3r}\sqrt{2r}\,\|R_0\|_F$$

(h) Let $K = R_0 + R_1$
Let A,B be two matrices, then we know that $rank(A + B) \leq rank(A) + rank(B)$
Proof:-
Let $x_1, x_2, ..., x_m$ be a basis of the rowspace of Matrix A and $y_1, y_2, ..., y_n$ be the basis of the rowspace of Matrix B. As we know every row of Matrix A+B can be written as a linear combination of $x_1, x_2, ..., x_m, y_1, y_2, ..., y_n$from this we can say that $rank(A + B) \leq rank(A) + rank(B)$
So using above theorem we can say that

$$rank(R_0 + R_1) \leq rank(R_0) + rank(R_1)$$

as $rank(R_0) \leq 2rank(X_0)$ and $rank(R_1) \leq 3r$, we can say that

$$rank(R_0 + R_1) \leq 2r + 3r$$

$$rank(R_0 + R_1) \leq 5r$$

(i) As we know $R = (R_0 + R_c) = R_0 + \Sigma_{i\geq 1} R_i$, we can write

$$R = (R_0 + R_1) + \Sigma_{i\geq 2} R_i$$

$$A(R) = A(R_0 + R_1) + A(\Sigma_{i\geq 2} R_i)$$

Now on applying triangle inequality we can say that

$$\|A(R)\| \geq \|A(R_0 + R_1)\| - \|A(\Sigma_{i\geq 2} R_i)\|$$

and applying triangle in equality for $\|\Sigma_{i\geq 2} A(R_i)\|$, will give

$$\|\Sigma_{i\geq 2} A(R_i)\| \leq \Sigma_{i\geq 2} \|A(R_i)\|$$

By using the above two equations we can say that

$$\|A(R)\| \geq \|A(R_0 + R_1)\| - \Sigma_{i\geq 2} \|A(R_i)\|$$

(j) As the rank of $R_0 + R_1 \leq 5r$, now on applying restricted isometric property on matrix $R_0 + R_1$, we will get

$$(1 - \delta_{5r}(A)) \|R_0 + R_1\|_F \leq \|A(R_0 + R_1)\| \leq (1 + \delta_{5r}(A)) \|R_0 + R_1\|_F$$

Similarly on applying restricted isometric property on matrix R_i for all $i \geq 2$(as rank of each matrix $rank(R_i) \leq 3r$) , we will get

$$(1 - \delta_{3r}(A)) \|R_i\|_F \leq \|A(R_i)\| \leq (1 + \delta_{3r}(A)) \|R_i\|_F$$

From this we can say that

$$\Sigma_{i\geq 2} \|A(R_i)\| \leq \Sigma_{i\geq 2}(1 + \delta_{3r}(A)) \|R_i\|_F$$
$$\Sigma_{i\geq 2} \|A(R_i)\| \leq (1 + \delta_{3r}(A))\Sigma_{i\geq 2} \|R_i\|_F$$

And from the first equation

$$(1 - \delta_{5r}(A)) \|R_0 + R_1\|_F \leq \|A(R_0 + R_1)\|$$

Using the above two equations we can write

$$\|A(R_0 + R_1)\| - \Sigma_{i\geq 2} \|A(R_i)\| \geq (1 - \delta_{5r}(A)) \|R_0 + R_1\|_F - (1 + \delta_{3r}(A))\Sigma_{i\geq 2} \|R_i\|_F$$

From this equation an the equation we proved in part (i) (i.e $\|A(R)\| \geq \|A(R_0 + R_1)\| - \Sigma_{i\geq 2} \|A(R_i)\|$), we can say that

$$\|A(R)\| \geq (1 - \delta_{5r}(A)) \|R_0 + R_1\|_F - (1 + \delta_{3r}(A))\Sigma_{i\geq 2} \|R_i\|_F$$

(k) We found X^* using $X^* = argmin_X \|X\|_*$ where A(X)=b, where $b = A(X_0)$
From this we can say that $AX^* = b$
Now,

$$A(X - X^*) = A(X) - A(X^*) = b - b = 0$$

So we can say that,

$$A(R) = A(X - X^*) = 0$$

(l) We can from equation 3.7 that,

$$\|A(R)\| \geq ((1 - \delta_{5r}(A)) - \frac{9}{11}(1 + \delta_{3r}(A))) \|R_0\|_F$$

Now as we saw in the part(k) that A(R)=0, and to show that $R_0 = 0$ which further implies $X^* = X_0$, we need the right hand factor to be greater than 0,i.e.

$$(1 - \delta_{5r}(A)) - \frac{9}{11}(1 + \delta_{3r}(A)) \geq 0$$
$$11(1 - \delta_{5r}(A)) \geq 9(1 + \delta_{3r}(A))$$
$$11 - 11\delta_{5r}(A) \geq 9 + 9\delta_{3r}(A)$$

On rearranging the above equation we will get

$$11\delta_{5r}(A) + 9\delta_{3r}(A) \leq 2$$

Question 4.

1. The theorems on low rank matrix completion require that the singular vectors be incoherent with the canonical basis due to the following reason. The rows of the sampling operator are dense in the canonical basis. That is, the sampling operator is a row-subsampled version of the identity matrix. If the singular vectors are also dense in the canonical basis, then there is a possibility that the singular vectors lie in the null space of the sampling operator. We cannot recover the matrix if the singular vectors lie in the null space of the sampling operator. This won't be the case if the singular vectors are *spread out*, i.e; they are incoherent with the canonical basis.

   Also, the paper's main result states that if a matrix has row and column spaces that are incoherent with the standard basis, then nuclear norm minimization can recover this matrix from a random sampling of a small number of entries.

2. The rank minimization problem mentioned in Equation (1.13) is

$$\text{minimize } rank(X)$$

$$\text{subject to } f_i^* X g_j = f_i^* M g_j,\ (i,j) \in \Omega$$

   where f_i and g_i are orthonormal bases in $\mathbb{R}^n$.
   To see that Theorem (1.3) works for the above case, we note that there exist unitary transformations F and G such that $e_j = Ff_j$ and $e_j = Gg_j$, for each $j \in [n]$. Hence,

$$f_i^* X g_j = e_i^*(FXG*)e_j$$

   Now, if the singular vectors of $FXG*$ are incoherent with the canonical basis, then we can find the unique solution by minimizing the nuclear norm. Therefore, for Theorem (1.3) (as given in the paper) to hold, we need the column and row spaces of M be respectively incoherent with the basis (f_i) and (g_i). This will ensure that the rows of M are not in the null space of g_i and the columns of M are not in the null space of f_1. Hence, the matrix M can be recovered.

Question 5.

Title:A Unified Approach to Salient Object Detection via Low Rank Matrix Recovery
Link: Here
Author: Xiaohui Shen and Ying Wu
Published Year: 2012
Venue : 2012 IEEE Conference on Computer Vision and Pattern Recognition

Aim In this paper, the authors propose a unified model to integrate bottom-up, lower-level features and top-down, higher-level priors for salient object detection. They represent an image as a low-rank matrix plus sparse noises in a learned feature space, where the low-rank matrix explains the non-salient regions (or background), and the sparse noises indicate the salient regions.

This is because the background usually lies in a low-dimensional subspace, while the salient regions that are different from the rest (i.e., deviating from this subspace) can be considered as noises.

Therefore, salient regions can be identified by performing low rank matrix recovery using the Robust PCA technique.

To ensure the model to be valid for visual saliency, a linear transformation of the feature space is introduced and learned. Higher-level priors can be naturally integrated into this model. Saliency is then jointly determined by low-level and high-level cues in a unified way. Higher level knowledge is then converted to pixel-wise priors and incorporated to this model to achieve better performance.

Problem being solved

The paper proposes an approach for image saliency, the process of applying image processing and computer vision algorithms to automatically locate the most "salient" regions of an image. While the salient regions are mostly unique, the inverse might not necessarily be true. Thus, to differentiate real salient regions from other unique/high contrast parts, priors from higher-level knowledge need to be integrated.

Objective Function

The paper attempts to solve the problem using low rank matrix recovery. The main objective function is given as,

$$(L^*, S^*) = arg\min_{\text{L,S}}(\|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1)$$

Such that $\mathbf{TFP} = \mathbf{L} + \mathbf{S}$
Here $\|L\|_*$ is the nuclear norm of L, and $\|.\|_1$ indicate $l_1$norm.
This equation is very similar to the cost minimization function in low rank matrix recovery. Here, L has low rank and S is sparse.

- Given an image, we extract different types of visual features around each pixel, including:Color filter, Steerable pyramids filters, Gabor filters,...etc. we use a total of 53 filters.
- 'D' is the dimension of the feature vector (D = 53 here).
- 'N' is the number of segments.(i.e. After feature extraction, we perform image segmentation based on the extracted features by mean-shift clustering. We select spatial and feature bandwidths to over-segment the image so that the background also contains multiple segments even if it is visually homogeneous.)
- $P = diag(p_1, p_2, ..., p_N), P \in R^{N*N}$ where each p_i is a prior probability (i.e probability of being salient for each segment).
- 'L' is the low-rank matrix corresponding to the background. $L \in R^{D*N}$
- 'S' is a sparse matrix representing the salient regions. $S \in R^{D*N}$
- $F = [f_1, f_2, ..., f_N].F \in R^{D*N}$, each f_i = feature vector of size R^{D*1}
- We learn a linear transformation T from a set of training images. After transformation, features can be represented as $g_i = Tf_i$ where $T \in R^{DxD}$. Accordingly, $G = TF = [g_1, g_2, ..., g_N]$

Note:- Matrix T is learned in the following way

$$T^* = arg\min_{\mathbf{T}} O(T) = \frac{1}{K}\Sigma_{k=1}^{K} \left\| TF_kQ_k \right\|_* - \gamma \left\| T \right\|_*$$

- $Q_k = diag(q_1, q_2, ...q_N)$, $Q_k \in R^{N*N}$.Q_k is the saliency indicator of images.F_k is the feature representation.