Flash Matting

Team FFT:


```
Aryan Kharbanda (2019101018)     Pranav Kannan(2019111033)
Kushal Jain (2019111001)         Tathagato Roy(2019111020)
```


Motivation

Extracting Foreground objects from Images have various applications especially in the making Digital Media.The most common method to achieve this by using Blue/Green background(Chroma - Keying).But this method is extremely expensive and requires well-controlled studio environment.Hence we aim to achieve similar results by leveraging pair of flash/no flash image.

Introduction

Formally a Image can be modelled using the following equation

$$I = \alpha * F + (1 - \alpha) * B$$

$$\text{Where I is the Image ,F is the the foreground image and B is the background Image}$$

The objective is to compute F,B and α from the observation I .This is underconstrained problem as we have to compute 7 values from 3 known values.Hence we utilize flash image pairs to make our problem more feasable.

The primary Assumption of our formulation is that the in the flash and no flash pair the background is at sufficient distance from the foreground and as such the effect of the flash in the background is negligible compared to the foreground object.
Mathematically we have :

$$I = \alpha * F + (1 - \alpha) * B$$
$$I^f = \alpha * F^f + (1 - \alpha) * B$$

Where I^f is the flash image and F to the power f is the flash foreground object and the rest is same as defined above.

Note the Background is assumed to be the same in both case

We define the flash only image as :

$$I' = I^f - I = \alpha(F^f - F) = \alpha F'$$

Where $F' = F^f - F$

We can estimate F' and α using the Bayesian formulation described in [2] and use the α value to recover F and B.

The Bayesian Formulation known as Bayesian Matting will be described now.Before that we need to describe what is a trimap.
A trimap of any image is an image of the same size as the input image which divides the pixels of the input image into 3 regions.The 3 regions are classified as "Surely Foreground" and "Surely Background" and "Unknown Region".
The quality of the trimap is a strong determiner in the result obtained.We will describe procedure of obtaining a trimap.For now we will assume the trimap is obtained.

Bayesian Matting of Flash only Image(Foreground Flash Matting)

For each unknown pixel we try to estimate F' and α given the flash only image F'

We want to maximize the following Probability :

$$P(F', \alpha | I')$$

Therefore :

$$\underset{F', \alpha}{\operatorname{argmax}} P(F', \alpha | I') = \frac{P(I' | F', \alpha) * P(F') * P(\alpha)}{P(I')}$$

Bayes Theorem

P(I') is a constant and can be ignored and Maximizing a product is more difficult than sums hence we introduce log likelyhood and our final result reduces to :

$$\underset{F',\alpha}{\operatorname{argmax}}\{L(F',\alpha|I') = L(I'|F',\alpha) + L(F') + L(\alpha)\}$$

The exact method to solve this particular optimization problem can be learnt from [2] , but we use a very similar method to solved the Joint Bayesian Flash Matting equation where we provide a more detailed descrption of our methods.

Joint Bayesian Flash Matting

We want to estimate F,F',B,α from observations I,I' for every unknow pixel.
Using the Log Likelyhood formulation described above we get :

$$\underset{F',F,B,\alpha}{\operatorname{argmax}}\{L(F,F',B,\alpha|I,I') = L(I',I|F,B,F',\alpha)\}$$

$$\implies \underset{F',F,B,\alpha}{\operatorname{argmax}}\{L(I|B,F,\alpha) + L(I'|\alpha,F') + L(F) + L(B) + L(F') + L(\alpha)\}$$

L(I) and L(I') are constant and ignored from the above equation
L(α) is assumed to be constant as there is no good prior for α and hence removed from maximization problem

The first two term measure the fitness of the solved variable with respect to the Matting Equations.
They are approximated using a gaussian distribution using the below idea.

$$P(I|B,F,\alpha) = e^{\frac{-||I-\alpha*F-(1-\alpha)*B||}{\sigma_I^2}}$$

$$\implies L(I|B,F,\alpha) = -\frac{||I-\alpha*F-(1-\alpha)*B||}{\sigma_I^2}$$

$$(and)$$

$$P(I'|F',\alpha) = e^{\frac{-||I'-\alpha*F'||/}{sigma_I^2}}$$

$$\implies L(I'|F',\alpha) = -\frac{||I'-\alpha*F||}{\sigma_I^2}$$

The F,F' and B is estimated by looking at the neighbourhood of the pixel and analysing the known pixel intensities to compute a statistical distribution.The known pixels are clustered and for each cluster, the mean $\bar F$ and Covariance Σ_F is computed. The mean approximation along with the Convariance is used to compute Log Likelyhood terms.

$$L(F) = (F - \bar{F})^T \Sigma_F^{-1} (F - \bar{F})$$
$$L(B) = (B - \bar{B})^T \Sigma_B^{-1} (B - \bar{B})$$
$$L(F') = (F' - \bar{F'})^T \Sigma_F'^{-1} (F' - \bar{F'})$$

We then get the whole Maximization Equation.Partially differentiating it with respect to F,F' and B we get following Linear System of Equations which we need to solve

$$A = \begin{bmatrix} \Sigma_F^{-1} + \mathcal{I}\{\frac{\alpha}{\sigma_I}\}^2 & \mathcal{I}\alpha(1-\alpha)\sigma_I & 0 \\ \mathcal{I}\alpha(1-\alpha)\sigma_I & \Sigma_B^{-1} + \mathcal{I}\{\frac{\alpha}{\sigma_I}\}^2 & 0 \\ 0 & 0 & \Sigma F'^{-1} + \mathcal{I}\{\frac{\alpha}{\sigma_I}\}^2 \end{bmatrix}$$

$$X = \begin{bmatrix} F \\ B \\ F' \end{bmatrix}$$

$$B = \begin{bmatrix} \Sigma_F^{-1}\bar{F} + I\frac{\alpha}{\sigma_I^2} \\ \Sigma_B^{-1}\bar{B} + I\frac{(1-\alpha)}{\sigma_I^2} \\ \Sigma F'^{-1}\bar{F'} + I'\frac{\alpha}{\sigma_I^2} \end{bmatrix}$$

Where $\mathcal{I}$ is a 3X3 Identity matrix and 0 is 3X3 Zero Matrix and rest are same as defined below

The Linear Equation is

$$AX = B$$

We get another Equation :

$$\alpha = \frac{\sigma_I^2 (F-B)^T (I-B) + \sigma_I^2 F'^T I'}{\sigma_I^2 (F-B)^T (F-B) + F'^T F'}$$

Solving the above two equations we get the values that maximizes the Log Likelyhood Optimisation Problem

Trimap Generation

Given an image, digital matting is extracting a foreground element from the background. Standard methods are initialized
with a trimap, a partition of the image into three regions: a definite
foreground, a definite background, and a blended region where pixels
are considered as a mixture of foreground and background colors also known as unknown regions. Recovering these colors and the proportion of mixture between both is an under-constrained inverse problem, sensitive to its initialization: one has
to specify an accurate trimap, leaving undetermined as few pixels as possible.

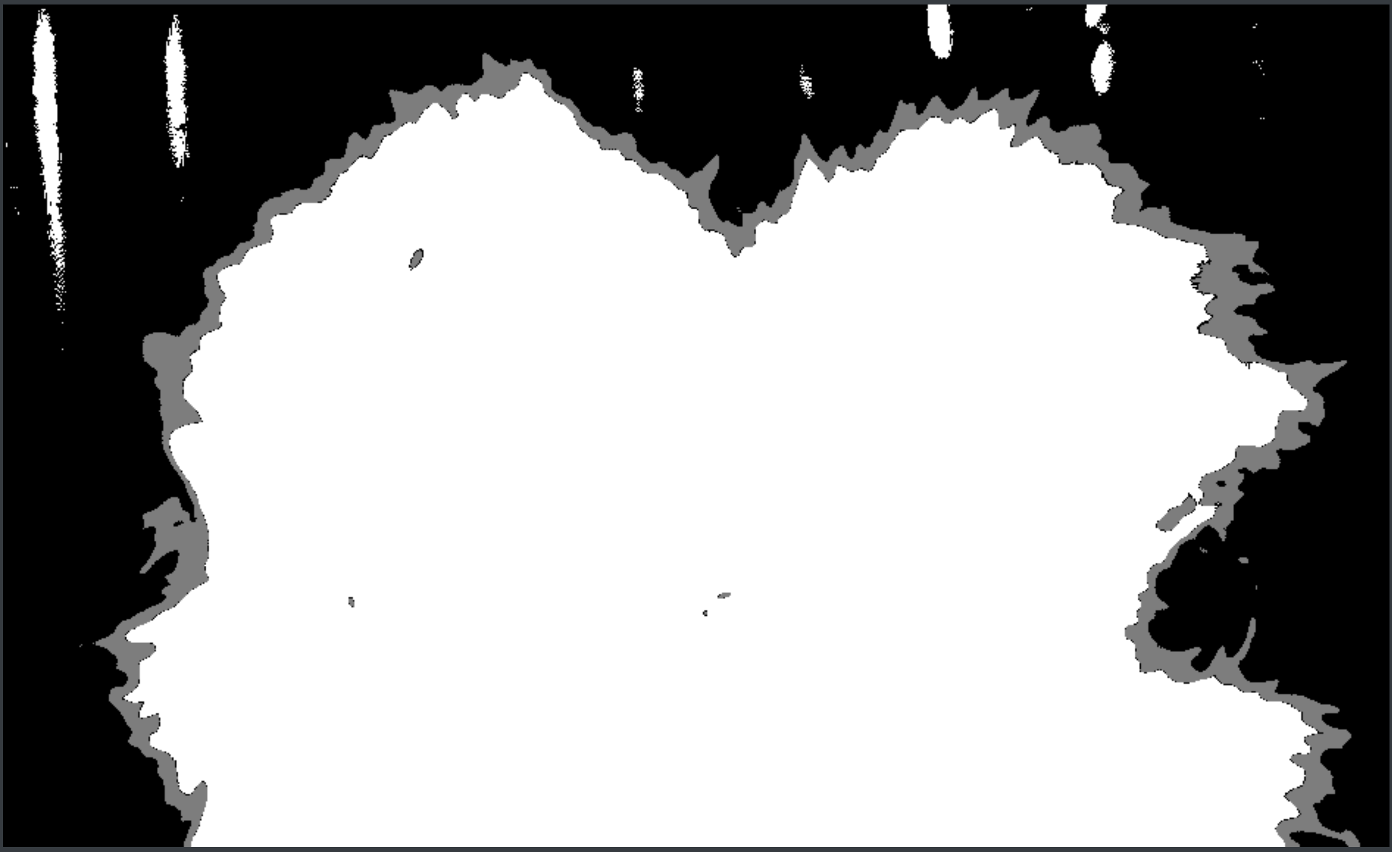
We estimate the trimap from I' using a technique similar to Canny's two-pass method [1986] for edge detection. In the first pass, a global high threshold T is used to detect a foreground region with high confidence. We set T as the first local minimum of the histogram (128 bins) of the intensity of I'. The histogram is smoothed using a Gaussian kernel (with a sigma of 5) to reduce noise. In the second pass, a lower threshold of 0.5T is used to detect foreground regions with lower confidence.

RESULT

No Flash

Flash

Trimap

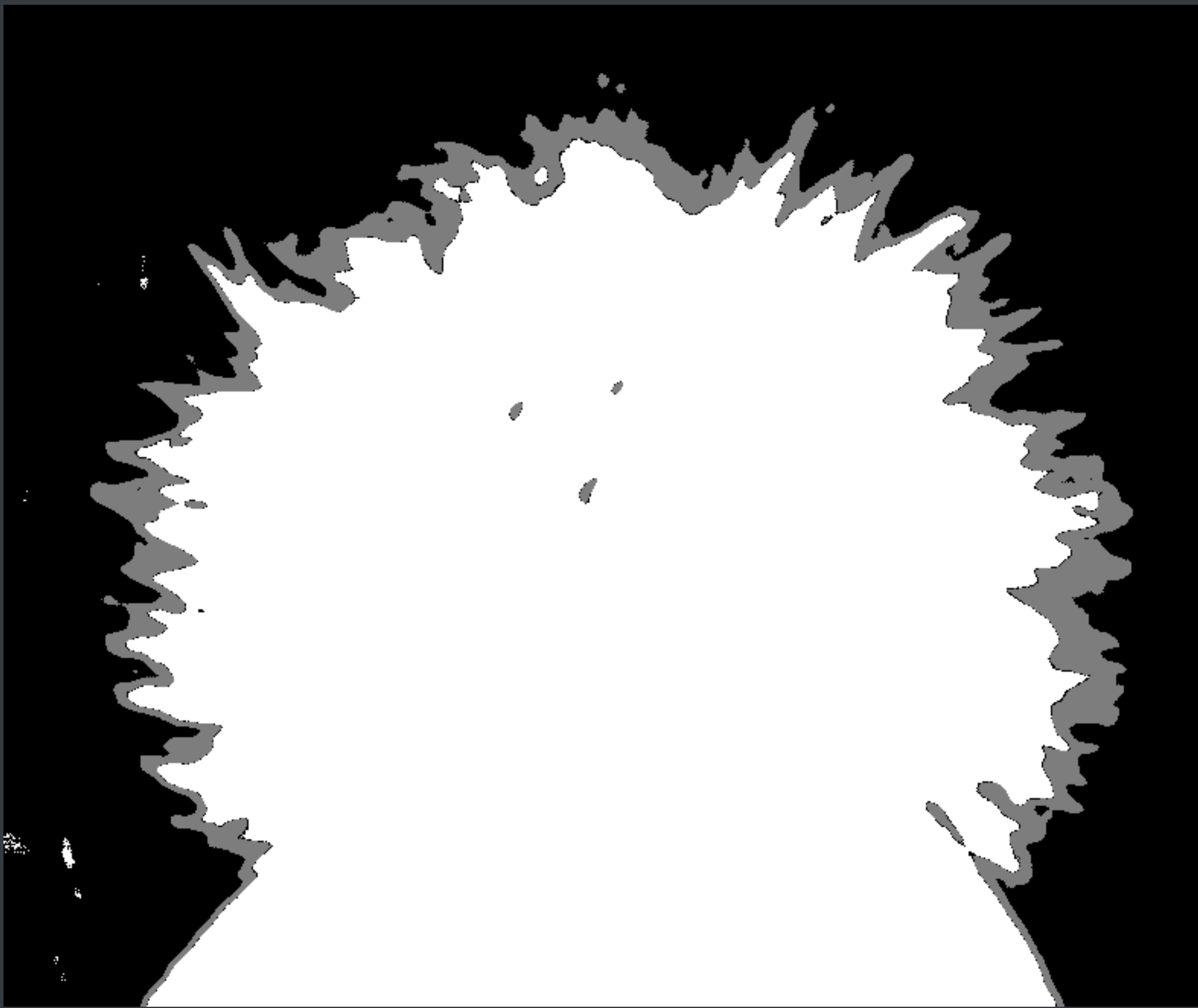


Foreground Image

No Flash

Flash

Trimap



Flash Matte (Foregroun)

Flash

No Flash

Trimap

Flash Matte

REFERENCES

[1] Sun, J., Li, Y., Kang, S. B., & Shum, H.-Y. (2006). Flash matting. In ACM Transactions on Graphics (Vol. 25, Issue 3, pp. 772–778). Association for Computing Machinery (ACM).

[2] Yung-Yu Chuang, Brian Curless, David H. Salesin, and Richard Szeliski. A Bayesian Approach to Digital Matting. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR 2001)*, Vol. II, 264-271, December 2001

[3] Venkatesh Praveen , Bayesian-Matting , (2021), GitHub repository https://github.com/praveenVnktsh/Bayesian-Matting

[4] Ya\ug\iz Aksoy, Changil Kim, Petr Kellnhofer, Sylvain Paris, Mohamed Elgharib, Marc Pollefeys, & Wojciech Matusik (2018). A Dataset of Flash and Ambient Illumination Pairs from the Crowd. In Proc. ECCV.