Image matting by Bayesian approach

May 22, 2022

P.Rajesh, EE21M019, M.Tech SPCOM
Guided by **Dr.Subrahmanyam Gorthi**

1 Abstract

This paper proposes a new Bayesian framework for solving the matting problem, i.e. extracting a foreground element from a background image by estimating an opacity for each pixel of the foreground element. Our approach models both the foreground and background color distributions with spatially varying sets of Gaussians, and assumes a fractional blending of the foreground and background colors to produce the final output. It then uses a maximum-likelihood criterion to estimate the optimal opacity, foreground and background simultaneously. In addition to providing a principled approach to the matting problem, our algorithm effectively handles objects with intricate boundaries, such as hair strands and fur, and provides an improvement over existing techniques for these difficult cases.

2 Introduction

In digital matting, a foreground element is extracted from a background image by estimating a color and opacity for the foreground element at each pixel. The opacity value at each pixel is typically called its alpha, and the opacity image, taken as a whole, is referred to as the alpha matte or key.

One common approach is to use a background image of known color (typically blue or green) and make certain assumptions about the colors in the foreground (such as the relative proportions of red, green, and blue at each pixel); these assumptions can then be tuned by a human operator.

Other approaches attempt to pull mattes from natural (arbitrary) backgrounds, using statistics of known regions of fore ground or background in order to estimate the foreground and background colors along the boundary. Once these colors are known, the opacity value is uniquely determined.

The most common compositing operation is the over operation, which is summarized by the compositing equation:

$$C = \alpha F + (1 - \alpha)B \tag{1}$$

where C, and B are the pixel's composite, foreground, C, B and background colors, respectively, and is the pixel's opacity component used to linearly blend between foreground and background.

Blue screen matting was among the first techniques used for live action matting. The principle is to photograph the subject against a constant-colored background, and extract fore ground and alpha treating each frame in isolation. This single image approach is underconstrained since, at each pixel, we have three observations and four unknowns.

3 Bayesian Framework

We know the observed colour as the pixel value in the image. Let us denote this pixel as C. We don't know the colour of the foreground, or the background, and hence need to estimate these by using a window around the area of interest. We can represent this mathematically as maximizing the probability of particular values of F, B and a given the observed colour C as arg maxF,B.c, P(F, B, a C). We need to find the values of F, B and a for all pixels that are marked unknown in the trimap.
Using the Bayes theorem and assuming that the foregroung pixels and the background pixels are independent, we can rewrite this as:

$$argmax_{F,B,\alpha} P(F,B,\alpha|C) = argmax_{F,B,\alpha} \frac{P(F,B,\alpha).P(F).P(B).P(\alpha)}{P(C)} \tag{2}$$

Since a product is more expensive to compute than addition, and there are several other operations that may involve exponentiation down the lane, we shall convert this into logarithmic scale. Let us represent log P L. This gives:

$$argmax_{F,B,\alpha} L(C|F,B,\alpha) + L(F) + L(B) + L(\alpha) \tag{3}$$

Let us now define the above terms in the equation: o In this method, the term L(a) is assumed to be a constant, and is eliminated from the maximization equation. o Since the term L(C F, B, a) gives a likelihood of C given the values F, B and a, we can say that a Gaussian distribution centred at the observed colour can model the likelihood. We can hence represent this term as:

$$L(C|F,B,\alpha) = \frac{-||C - \alpha F - (1-\alpha)B||^2}{\sigma_C^2} \tag{4}$$

where σ_C^2 is the variance in the image

We use data near the region of interest to estimate the terms L(F) and L(B). First, we take a small window of interest near the pixel marked unknown. Let the window size be N x N x 3 (3 for the 3 channel image BGR). Now, in order to more accurately model the distribution, let us weight the contributions of the pixels before proceeding with the math.

The weights can be determined by defining a Gaussian kernel multiplied with the pixel opacity squared. i.e., $w_i = \alpha_i^2 g_i \forall_i \in$ window. For the background, we can use a weighting window defined by $w_i = (1 - \alpha_i^2) g_i \forall_i \in$ window. The Guassian kernel helps emphasize colours that are closer to the unknown region than those that are farther away since the colours closer will represent a better estimate of the actual foreground or background due to the continuous nature of the image (excluding noise).

Once we get the weighting window, we now cluster the entire window into some n clusters. In my implementation, I have used the KMeans clustering algorithm with n = 5. o We will be fitting the foreground and background into a Gaussian distribution, we will need both the covariance and the means of these clusters.

with these n clusters, we find the weighted mean color of each cluster

$$\overline{F} = \frac{1}{W} \sum_{i \in N} w_i F_i \tag{5}$$

Now, we find the weighted covariance between the RGB channels of the particular cluster. This will help establishing the dependence of each of the 3 channels with each other.

$$\sum_{F} = \frac{1}{W} w_i (F_i - \overline{F})(F_i - \overline{F})^T \tag{6}$$

Now that we have the mean and covariance ,let us fit a Gaussian distribution onto our foreground.we can hence represent the likelihood of foreground of a particular cluster as:

$$L(F) = -(F - \overline{F})^T \sum_{F} {}_{-1} (F - \overline{F})/2 \tag{7}$$

By representing a similar case of background , we can find the value of L(B) as well.

Now that we have all the terms in the maximization equation , we take the derivative of the equation with respect to F and B and set them to zero , assuming that α is constant .By doing this ,we get an equation:

$$\begin{bmatrix} \Sigma_F^{-1} + I\alpha^2/\sigma_C^2 & I\alpha(1-\alpha)/\sigma_C^2 \\ I\alpha(1-\alpha)/\sigma_C^2 & \Sigma_B^{-1} + I(1-\alpha)^2/\sigma_C^2 \end{bmatrix} \begin{bmatrix} F \\ B \end{bmatrix} = \begin{bmatrix} \Sigma_F^{-1}\bar{F} + C\alpha/\sigma_C^2 \\ \Sigma_B^{-1}\bar{B} + C(1-\alpha)/\sigma_C^2 \end{bmatrix}$$

This can be solved very easily since all of the values are known except alpha!

After obtaining the values of F,B, we substitute the constant value of F and B in the maximisation equation and arrive to a close form solution for the opacity as:

$$\alpha = \frac{(C - B).(F - B)}{||F - B||^2} \tag{8}$$

we are left two equations that can now be iteratively solved in order to reach a good solution that maximizes the likelihood equation

since there were several clusters that we found previously ,and we solved for only one cluster ,we iterate through pairs of the foreground and background clusters to see which ordered pair gives the maximum likelihood for the probability.

4 Results

Given Image

background image