Semi-automated binary segmentation with 3D MRFs

Submission for Assignment 1 EE55C1

Liam Byrne
Student ID: 18326579
Trinity College Dublin

***Abstract*—This document contains a solution for the problem of extracting foreground elements from a background image, known as matting. The approach taken, applies Bayes' theorem, specifically Maximum a posteriori (MAP) estimation, which uses a likelihood function based on maximum-likelihood criterion and a prior probability function based on Markov random field to predict foreground and background pixels at the same time.**

I. Introduction

Matting allows for different foreground and background elements of an image to be removed This allows elements to be composited to create special effects, change the background of an image, or remove unwanted objects from a scene. The compositing processed can be expressed as:

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

F is the foreground image, B is the background image, and α is the transparency or opacity of the foreground image. Y, is a composite of the two images, where the foreground image appears on top of the background image.

Composites were originally produced by manually overlaying image cutouts. Compositing technology has improved greatly thanks to computers however producing composites remains a laborious process if done frame by frame on video. In this document the compositing process is undertaken on video. Color keying makes compositing process quicker by shooting the subject against a plain, uniformly colored background which can be removed and replaced with another background. This technique is ideal for relatively simple and quick binary matting which separates the foreground object from the background by creating a binary mask, where the foreground pixels are marked as white (1) and the background pixels are marked as black (0).

Binary matting can be accomplished using various techniques, one of which is Bayesian matting. In Bayesian matting, the probability of each pixel being part of the foreground or background (α) is calculated using Bayesian inference and maximum likelihood estimation. Maximum likelihood approach uses Gaussian distributions to model the color variations in the foreground and background regions.

II. Modeling Mathematically

A. Bayes Theorem

In order to perform binary segmentation we must discover the binary mat value alpha α at each pixel location $x = [h, k]$ in image I of frame n based on its colour value $I(x)$. If the probability of α is 1 (i.e. foreground) is greater than the probability α is 0 (i.e. background) then the pixel is set to 1 for foreground and vice versa. We can calculate $P(\alpha(x) \mid I(x))$ the posterior probability of the opacity $\alpha(x)$ at pixel x given the observed data $I(x)$ using Bayes theorem.

$$P(\alpha(x) \mid I(x)) = \frac{p(I(x) \mid \alpha(x)) \cdot p(\alpha(x))}{p(I(x))} \tag{2}$$

The numerator $p(I(x) \mid \alpha(x)) \cdot p(\alpha(x))$ represents the joint probability of the data and the opacity, where $p(I(x) \mid \alpha(x))$ is the likelihood of the observed data given the opacity, and $p(\alpha(x))$ is the prior probability of the opacity. The denominator $p(I(x))$ is the marginal likelihood of the data i.e. the total probability of observing the data $I(x)$.

The likelihood will be obtained using Maximum a posteriori (MAP) estimation with Maximum Likelihood criterion and Markov random field theory will be implemented as the prior probability function.

B. Maximum Likelihood Estimation

A Gaussian distribution of background pixels will help us tell whether a pixel belongs to the foreground or background. The likelihood function for a background pixel can be written as:

$$P(I|\alpha = 0) = \frac{1}{\sqrt{2\pi\sigma_B^2}} e^{-\frac{(X-\mu_B)^2}{2\sigma_B^2}} \tag{3}$$

where B represents the background Gaussian Distribution for frame n. μ_B and σ_B^2 are the mean and variance of the background pixel values.

The Gaussian distribution of the background is used as it is assumed the subject has been captured in front of a green screen meaning the background will be relatively consistent. Taking a sample of the background we can estimate the mean and variance of background pixels as well as average background pixel colour values. Assuming that the image is expressed in the YUV color-space channels we can express the probability distribution of the background with respect to the Gaussian distribution as

$$p(I|\alpha = 0)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-} \left(\frac{(Iy - By)^2}{2\sigma_y^2} - \frac{(Iu - Bu)^2}{2\sigma_u^2} - \frac{(Iv - Bv)^2}{2\sigma_v^2} \right) \tag{4}$$

For binary matting the pixel must be classified as foreground or background and for this we must set a threshold value. If the energy of the pixel in question crosses the threshold for being considered background it will be classified as background otherwise it is classified as a foreground pixel.

Rearranging the formula we can achieve a condition which if met determines that the pixel I belongs to the background.

$$\frac{(I_y - B_y)^2}{2\sigma_y^2} + \frac{(I_u - B_u)^2}{2\sigma_u^2} + \frac{(I_v - B_v)^2}{2\sigma_v^2} < \ln(k) \tag{5}$$

C. Markov Random Field

Obtaining a prior distribution can be challenging in Bayesian matting because it can require subjective input from the user. The choice of prior distribution can have a significant impact on the results of Bayesian estimation. Markov Random fields (MRFs) provide a method for systematically analysing frames of a video. MRF theory when applied to pixels assumes that neighboring pixels in an image are correlated, and these correlations can be modeled using the conditional dependencies between them. in this way MRFs can be used to segment an image into different regions by estimating the probability of a particular pixel belonging to a certain class or category. However, pixels in an image are generally highly correlated with each other which makes it difficult to specify an MRF using only the conditional probability structure. To model dependencies between neighboring pixels, a Gibbs distribution is used to specify the joint probability distribution of the pixel labels in an MRF. The energies $E_s(0)$ and $E_s(1)$ are calculated at all pixel locations x where V represents smoothness weight and q_k represents the neighbourhood of pixel values.

$$E_s(0) = \sum_{k=1}^{4} V(0, \alpha(x + q_k)) \tag{6}$$

$$E_s(1) = \sum_{k=1}^{4} V(1, \alpha(x + q_k)) \tag{7}$$

D. Maximum a Posteriori Estimation (MAP)

We can find the posterior distribution by combining the probability distribution i.e. the Maximum likelihood estimation, and the prior distribution i.e. the Gibbs distribution of pixel labels in an MRF.

$$E(0) = \ln(z) + [(I - I_\alpha)^T R_\alpha^{-1} (I - I_\alpha)] + \sum_{k=1}^{4} V(0, (\alpha(x + q_k))) \tag{8}$$

$$E(1) = \ln(z) + E(t) + \sum_{k=1}^{4} V(0, (\alpha(x + q_k))) \tag{9}$$

III. Algorithm and Optimisation

A. Algorithm: Maximum A posteriori

1) Measure Gaussian parameters for background somewhere in the example image. Get the mean and variance of each colour component in the background B_μ, σ_y^2, etc.
2) Set some threshold E_t for deciding a site is foreground.
3) At every pixel site
   a) measure $E(l)$, $E_s(0)$ (eqn 6) and $E_s(1)$ (eqn 7), where:
   $$E(l) = \frac{(I_y - B_y)^2}{2\sigma_y^2} + \frac{(I_u - B_u)^2}{2\sigma_u^2} + \frac{(I_v - B_v)^2}{2\sigma_v^2} \tag{10}$$
   b) Set $E(0) = E(l) + E_s(0)$ and $E(1) = E_s(l) + E_s(1)$
   c) If $E(0) < E(1)$, set $\alpha(x) = 0$, Otherwise $\alpha(x) = 1$

In the MAP algorithm provided above $E(l)$ represents the maximum likelihood energy function $E_s(0)$ and $E_s(l)$ are the MRF energy functions

B. Motion estimation

Frames in a video sequence often stay relatively consistent (with the exception of cut scenes). In segmentation tasks performed on video we can often use information from a pixels neighbourhood in the previous frame. 3D MRFs work on the assumption that the pixel neighbourhood in the current frame depend on the previous frame neighbourhood pixels. Motion estimation algorithms enable the transformations objects undergo between video frames to be estimated with motion vectors. We can use motion Vectors to allegorically predict the next frame based off of object movement. motion vectors we can track objects and pixels between frames. By accounting for motion we can provide the 3D MRF a more useful previous frame.

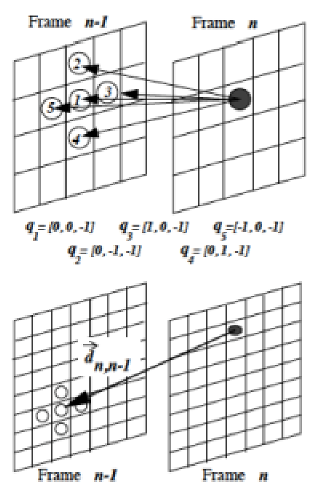


Fig. 1. MRFs for Video

IV. Nuke Implementation

The implementation was done in nuke using a combination of preset nodes and custom Blink-script nodes. The implementation was done in four different stages, first was the maximum likelihood estimation, second the 2-Dimensional Markov

Random Field, the spatiotemporal 3-Dimensional MRF and finally the Motion compensated 3D MRF. We ensured that the colourspace for the input frame sequence was in sRGB. The mean background values for each colour channel was known in advance an expression node was created that implemented part of eqn 10.

The prior was calculated initially using a neighbourhood of 8 pixels in the nodes encapsulated in the 2D MRF backdrop. Gradually increasing in complexity we implemented a 3series of 3D MRF Blinkscript node which took into account the pixels neighbourhood in the previous frame. Then used a vector generation node for estimating motion. The motion compensation was conducted by the I distort node which was applied to the image input and motion generated by the vector generation node. Similarly to the regular 3D MRF a timeoffset node was used with motion compensation to ensure the previous frame was being used to generate results.

V. Experiments and Performance Measures

VI. Results and Discussion

A. Maximum likelihood estimation

The maximum likelihood estimation resulted in errors both in the background section of the matte and in the foreground additionally there was a significant portion of the bottom left hand side of the matte that was incorrectly labbeled foreground which was the case for all of the algorithms. When the matte is applied to the input image we can see the green background is still clearly visible around the edges of the hair.

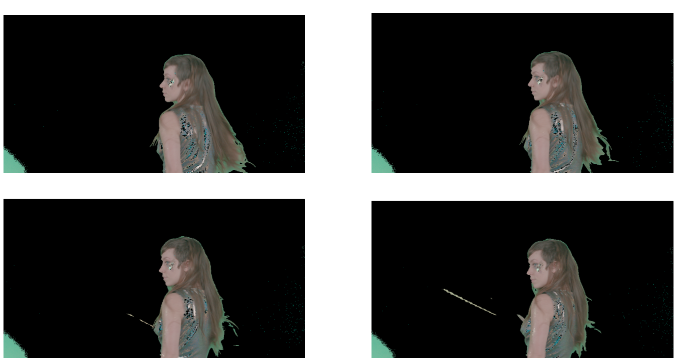
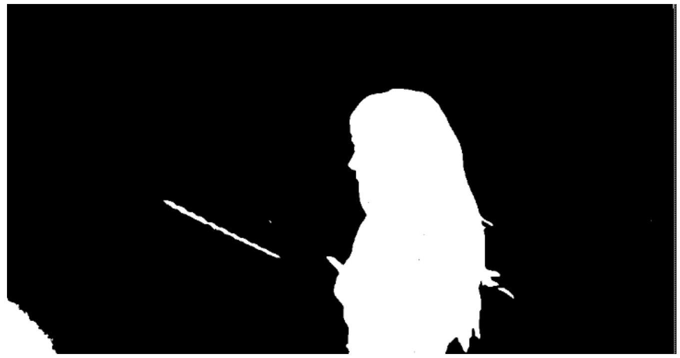


Fig. 2. Results of Maximum Likelihood Estimation

B. 2D MRF



The 2D MRF eliminated all of the background pixels incorrectly labelled foreground in the subject as the background pixels labelled foreground on the right hand side of the matte. There was still a significant amount of the green-screen showing through the edges of the hair making the matte unusable at any professional level.

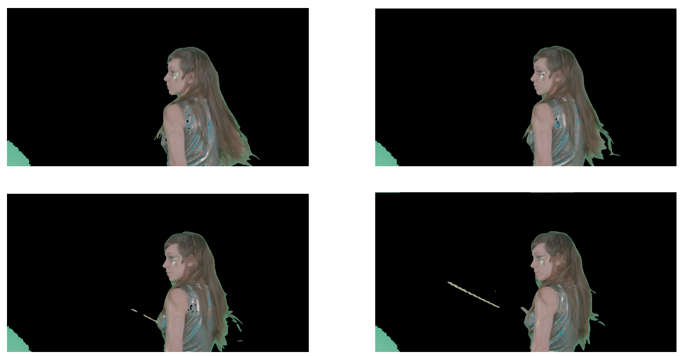


Fig. 3. 2D MRF

C. 3D MRF

The 3D MRF showed no significant improvement to the binary matte than was achieved with the 2D MRF. Both proceses showed very similar results with no black pixels showing through in the foreground however, there remained an unsatasfactory level of green background showing through the hair of the subject.

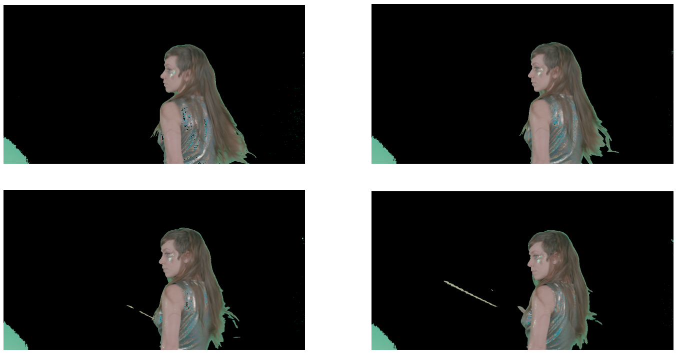


Fig. 4. 3D MRF

D. 3D MRF with Motion Compensation

The 3D MRF which included motion compensation performed similarly to the previous two methods, however, it managed to scheive a marginally better matte around the hair. overall this slight improvement in performance was not enough for a satisfactory composite image. The slight improvement around hair also came at the cost of pixels from the foreground being incorrectly labelled as black. Overall the matte acheived from implementing 3D motion compensated MRFs was more detailed than the relatively smooth mattes generated from the 2D and 3D MRFs.

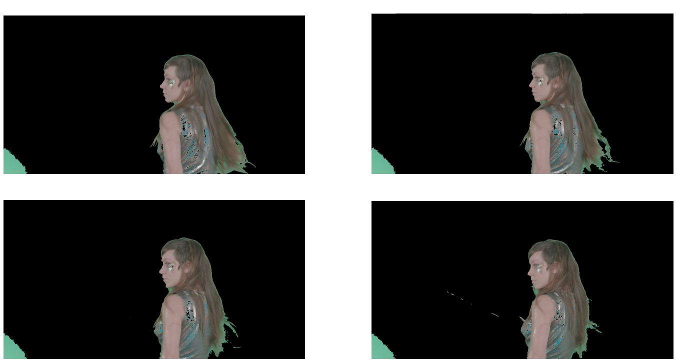
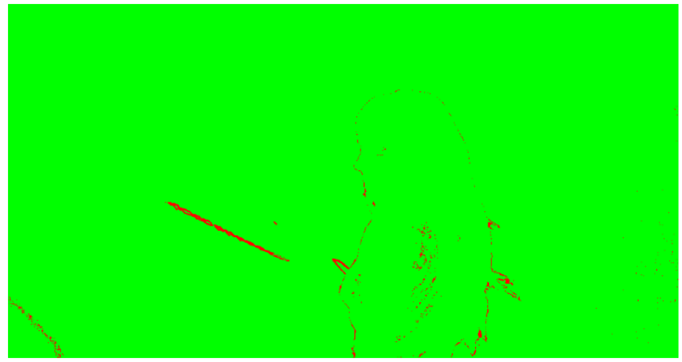


Fig. 5. Motion Compensated 3D MRF



VII. Comparison of 2D MRF after 2 iterations and Motion Compensated 3D MRF

The maximum likelihood estimation matte resulted in errors no matter what threshold was used. The 2D Markov Random Field matte resulted in a matte which was not tight enough particularly around hair. The 3D Markov Random Field without motion compensation generated matte appeared the best matte. The 3D Markov Random Field with motion compensation resulted in the background showing through the matte where it was not desired. None of the mattes implemented here would be suitable for use in industry and other avenues such as deep learning and other solutions which utilise minimal guidance from an artist.

VIII. Declaration

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar. I certify that this submission is my own work.

References

Appendix