MATTING WITH A DEPTH MAP

F. Pitié and A. Kokaram

Sigmedia Research Group, Electronic and Electrical Engineering Dept.
Trinity College Dublin
fpitie@mee.tcd.ie

ABSTRACT

Depth maps are becoming a readily available commodity of the stereo pipeline. We propose to make use of this new free information to improve a key step of postproduction that is matting. We extend the work of Levin *et al* on closed form matting to introduce two new depth-aware techniques. First we explore how depth can be used as an extra channel in the matting process. Then we see how depth can be used as a diffusion guide for matting. Our results show that both techniques can reduce the amount of time needed to pull a matte.

Index Terms— Stereo, Matting, Post-Production

1. INTRODUCTION

With the recent rise of stereo filming, access to scene geometry has become increasingly simpler. Already, widely used postproduction software platforms, such as *Nuke* and its stereo software suite *Ocula 2.0*, offer to the artists tools to extract depth maps from stereo footage. In the same way that motion estimation is now ubiquitous in 2D postproduction tools, we want to find ways of using these available depth maps back into the 2D postproduction pipeline. In this paper, we propose two techniques to integrate the depth map into one of the key tools of post-production: matting.

Pulling a matte from a film or video sequence is the excercise of cutting out an object from its background by creating a transparency mask or α *matte*, that is non-zero in the region of the object and zero otherwise [1]. To help the algorithms, recent techniques ask the artist to scribble a *trimap*, defining foreground, background and the unknown region to be pulled.

Since the successful work of Chuang *et al.* [1] on Bayesian Matting in 2001, the notion of Matting as an inference problem has been explored by several authors (see in particular Poisson Matting [2] and Inference Matting [3]). Recently Levin *et al.* [4] have proposed a number of remarkable advances by finding a closed form solution to the matting problem. Their solution is fast and extremely robust, being able to generate convincing mattes. The unknown trimap regions can be very much larger than those tolerated in other algorithms. This is perhaps the most interesting of the solutions generated to date.

Matting with Depth. If we consider matting as a two steps process with segmentation and transparency estimation, it is clear that any information about the relative depth of an object will help the segmentation step. There has been a great deal of research on using depth as a clue for segmentation [5] and some authors have also proposed to jointly segment and estimate the depth map [6]. Our application is slightly different: we want to reuse available depth maps to bring elements of geometry into the solution for matting. The solution should be able to work we current disparity estimators, hence accounting for the kind of quality that is currently available.

In previous work on Depth assisted Matting [7] two different methods for incorporating depth information into existing algorithms was proposed. First, even though matting operate principally in colour space, the authors suggest using depth as a fourth colour channel to Bayesian Matting and modifying the algorithm to cope with this fourth dimension. They recommend a second method for the Poisson matting type approach which operates on each colour channel separately. In this case a confidence map is developed that combines the estimated values for alpha for each colour channel and the depth map in a non-linear fashion. This essentially has the same effect of ignoring the depth information in areas where α is somewhere between zero and one.

Contribution. Our attempt is to bring depth inside the closed form matting framework. In a first step we consider, as in [7], the effect of using the depth as an extra colour channel. In the second technique, we have chosen to introduce a diffusion process that relates the gradient of the matte to the gradient of the depth map. This makes sense because the depth map is a better indication about the interface between objects than the gradient of the image. We also lay out a number of important implementation details that help to achieve greater stability and speed while computing the solution of the closed form solution.

Organisation of the paper. We present the closed form matting framework in section 2 and then our two techniques in section 3. Implementation improvements are then presented in section 4. Results follow in section 5.

This work was funded by the FP7 EU project i3Dpost.

2. CLOSED FORM MATTING [4]

The matting problem can be expressed with the following *matting* equation:

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i \quad (1)$$

where I_i is the colour vector for pixel i, F_i and B_i the colour of the foreground and background and α_i the transparency value. In the remarkable work of Levin et al. [4], the matting equation 1 is rewritten to be linear in the unknowns, by replacing F and B with two new variables a and b as follows:

$$\alpha_i = a_i^T I_i + b \quad (2)$$

For gray scale images, this formulation is similar to the one of eq 1. Indeed if we set $a = 1/(F - B)$ and $b = -B/(F - B)$, then both equations are similar. The difference appears for colour images. Instead of 6 unknowns (3 for F, 3 for B), we now ends up with 4 unknows (3 for a and 1 for b). The practical implication is that the problem is more constrained. The drawback is that F and B must be estimated separately.

In their work they impose a simple prior on a, b that assumes that for each pixel j, a, b is constant over a 3x3 image patch w_j. The energy to minimize has the following form:

$$J(\alpha, a, b) = \sum_j \sum_{i \in w_j} (a_j^T I_i + b_j - \alpha_i)^2 + \epsilon a_j^2 \quad (3)$$

The extra term ϵa_j^2 controls the smoothness of the matte and corresponds to a prior that $a \sim \mathcal{N}(0, 1/\epsilon^2)$. In effect, faced with the problem of manipulating α, a, b, Levin et al have chosen to integrate a, b out of the problem and so generate a marginalised estimate of $\hat{\alpha}$. What makes this possible is that the model is linear in the parameters a, b. There are very interesting implications of this step [8], and this marginal estimate for α is certainly not the same as the MAP estimate that other authors have been attempting to generate.

Levin et al show how the marginal estimate can be generated by writing the linearised equations for the whole unknown region of the trimap and then a deterministic solution is had simply with a matrix inversion that is closed form. See [4] for a detailed treatment.

At the end, the *closed form solution* is as follows:

$$(L + \lambda D_s) = \lambda b_s \quad (4)$$

where $\lambda > 0$ is some large number, D_s a diagonal matrix whose elements are one for the pixels on the scribbles of the trimap and 0 elsewhere and b_s a vector containing the values of α on the scribbles. The sparse matrix L is the *matting Laplacian*, with entries $L(i, j)$ as follows:

$$\sum_{k|(i,j) \in w_k} \left(\delta_{i,j} - \frac{1}{|w_k|} (1 + (I_i - \mu_k) \left(R_k + \frac{\epsilon}{|w_k|} \right)^{-1} (I_j - \mu_k) \right) \quad (5)$$

where R_k is the covariance matrix for the patch w_k.

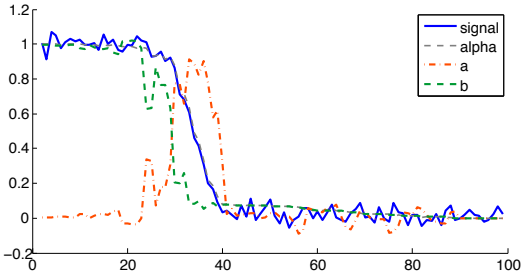

Fig. 1. Closed Form Matting on a 1D line. Outside the boundary (20-40), the mixing is turned off with $a \approx 0$.

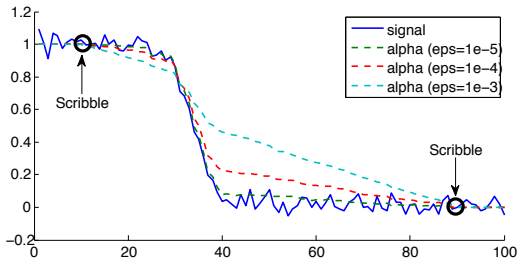

Fig. 2. Effect of the prior $p(a) = \mathcal{N}(0, 1/\epsilon)$ on the α matte for a 1D signal. Two scribbles have been assigned at position 10 and 90.

3. GUIDING THE MATTE

3.1. Depth Map as an Extra Channel

We first propose to see what happens when the depth is treated as an extra colour channel. This might seem a little strange as the depth is not a mixture of foreground and background depth layers but we show in Figure 1 that closed form matting essentially operates as a binary segmentation outside the object boundary. The figure reports the estimated value of α (gray), a (red) and b (green) for a simple 1D toy example with two scribbles. Note that outside the object boundary (ie. < 20 and > 40), the scaling factor a is almost null and the value of α is binary: either $\alpha = b = 0$ or $\alpha = b = 1$. Thus it is valid to use the depth as a channel for regions that are not on the boundary and in that sense it will help the matting.

To control the influence of the depth channel globally, we can use the prior penalty on the depth component of a, that we denote as a_D. The prior $\mathcal{N}(0, 1/\epsilon)$ on the components of a has two effects: 1) it a leverages the numerical inversion of the covariance matrix R_k in Eq. 5 and 2) it offers a smoothness on the α matte (see Figure 2). With four channels, the penalty in Eq (3) becomes $\epsilon(a_R^2 + a_G^2 + a_B^2) + \epsilon_D a_D^2$.

By increasing the inverse-variance ϵ_D of a_D, we can set $a_D = 0$ and switch off the influence of the depth channel. Choosing an appropriate value for ϵ_D allows us to mitigate the inaccuracy of the depth map and somehow get around the problem that the depth map is not a mixture of layer.

3.2. Depth Map as an Anisotropic Diffusion Map

We propose another tool to take advantage of the depth map. The idea is to add a diffusion process that favours constant values of α on areas where the depth is also constant. The penalty can be inserted inside the closed form matting framework as follows by finding the optimal α solution of:

$$\min_{\alpha} \sum_{j} \left(\min_{a,b} \sum_{i \in w_i} (a_j^T I_i + b - \alpha_j)^2 + \omega_{i,j}(\alpha_i - \alpha_j)^2 \right) + \epsilon a_i^2 \tag{6}$$

Typically the diffusion process will loosely follow the 3D topology: it is strong on uniform areas of the provided depth map, and can be stopped at depth discontinuities:

$$\omega(i,j) = \begin{cases} \mu \frac{1}{1+K||g_j - g_i||} & \text{if } i \in w_j \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

for some $K > 0$ and control parameter μ. With the scribbles constraint, we obtain the following linear system:

$$(L + \Omega + \lambda D_s)\alpha = \lambda b_s \tag{8}$$

where $\Omega = (\omega_{i,j})_{i,j \leq N}$ is a sparse matrix that drives the diffusion process.

Note that if the diffusion does not take into account the depth variations, i.e. $\omega_{i,j}$ is a constant, then the diffusion process is a ballooning energy that biases segmentation cuts to be at half-way between scribbles.

4. CONTRIBUTIONS TO THE IMPLEMENTATION

Numerical Stability. The proposed method follows the main steps of the closed-form solution as discussed in the Levin et al. paper [4], with our alterations for inclusion of the Depth information. Although the algorithm is mathematically well established, and theoretically stable, the implementation of the closed-form matting requires some care in the execution. The problem lies in the inversion of the sparse linear system of equation 5. The sparse matrix L is by construction symmetric positive definite, hence the inversion should be straightforward. In practice however, it is hard to enforce sufficient definiteness. The evaluation of the terms $d_{ij} = (I_i - \mu_k)^T R_k^{-1} (I_j - \mu_k)$ is the main issue. These terms can be undefined when the covariance matrix R_k becomes singular, which happens when the patch has a uniform colour. In Levin et al., it is proposed to use the inverse-variance ϵ as a way of increasing the definiteness of R_k. This is a standard alteration which results however in increasing the smoothness of α as seen previously. We propose instead to compute the SVD decomposition of the covariance matrix $R_k = U^T \Lambda U$ to invert the covariance matrix: $R_k^{-1} = U^T \Lambda^{-1} U$. Since the instability comes from the inversion of the small eigenvalues $(1/\lambda_k)$, we truncate small eigenvalues to some small value. The large eigenvalues are unchanged and the impact on the accuracy very small. With this modification, we can safely use very small values of for ϵ and obtain an α matte that follows more tightly the object boundary.

Solving the sparse linear system. We employ an iterative multi-scale technique for solving the sparse linear system. Namely we use the conjugate gradient with preconditioner to solve each level of the image pyramid. We chose the Incomplete Cholesky decomposition as a preconditioner, which we found to be quite efficient. On some pictures the incomplete Cholesky decomposition fails and we must then default to the slower incomplete LU preconditioner. Overall it takes around 1-10 seconds to pull a matte on a 2K plate on an intel Macbook 2008, using only one CPU thread. The computation time varies depending on the area of the picture to be processed and the success of the incomplete Cholesky preconditioner. In practice, the system is responsive enough to allow for interactive manipulation.

5. RESULTS

For all the results presented here the depth map has been generated using the stereo algorithm from Bleyer and Gelautz [9], which is fast enough to be practical in a post-production context. Resulting depth maps show significant amount of noise and do not follow object contours tightly. These are very typical of the kind of disparity map that artists would have at their disposition. Note that the occlusions are not marked, which will cause further problem for the segmentation at boundaries.

On Fig. 3 we show the effect of using the disparity map as diffusion process (middle row) and as an extra channel (bottom row). The original picture and its scribbles are on the top row. The diffusion is controlled by μ as presented in Eq. (7) and the influence of the disparity channel is controlled by ϵ_D as presented in section 3.1. Note that on the right column, where depth is the main element to guide matting, the imperfections of the depth map become visible in the pulled matte.

Comparison results with the closed form solution by Levin [4] are presented on Fig. 4. Note that it is hard to properly compare the original closed form matting to our extended the depth assisted matting. An artist can always draw more scribbles to fix the matting. A fair comparison is thus a difficult task. That said, we found that our tools offer useful extensions to the original solution. On the image sequence on Fig. 4, over 100 frames, an average of 8-10 scribbles per frame where required using our techniques, whereas an extra 5-6 scribbles where needed without the tools, which is a consequent amount of time saved.

6. DISCUSSION AND CONCLUSION

We have presented two techniques to introduce the depth information into the matting framework of Levin et al. [4]. Our approaches are 1) to use the depth as an extra channel and 2)

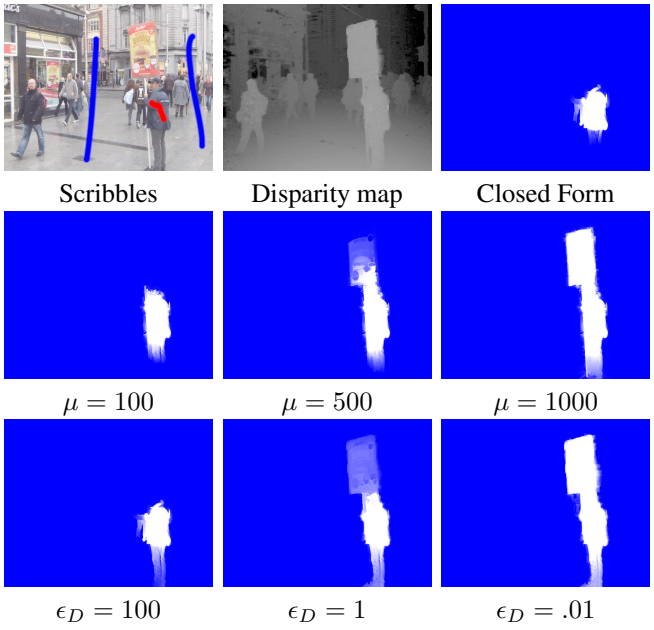

Fig. 3. Influence of using the disparity as: a diffusion process (middle row) and as an extra channel (bottom row). For the diffusion process, the input disparity map has been thresholded.

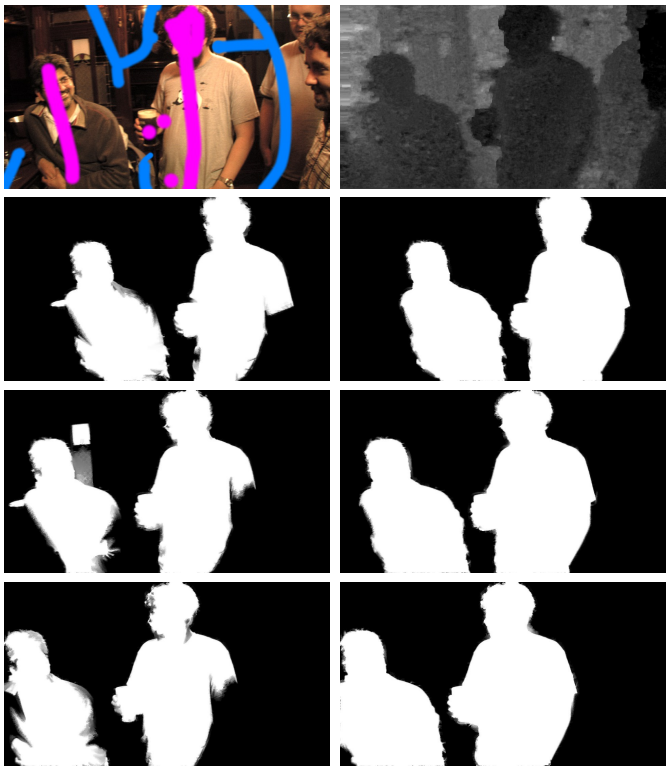

Fig. 4. On top: a frame picture with scribbles and its disparity map. Then on the left column: matting without using the depth. On the right: using our techniques.

to add an anisotropic diffusion process that biases alpha values to be smooth on areas of the picture that have uniform depth. Provided that the depth information gives a good clue, these tools can reduce the amount of scribbles necessary to pull a matte. Even though the quality of the results is dependent on the quality of the input depth map, it transpires that approximate depth maps can still lead to a noticeable help.

7. REFERENCES

[1] Yung-Yu Chuang, B. Curless, D.H. Salesin, and R. Szeliski, "A bayesian approach to digital matting," 2001, vol. 2, pp. II–264 – II–271 vol.2.

[2] Jian Sun, Jiaya Jia, Chi-Keung Tang, and Heung-Yeung Shum, "Poisson matting," in *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, New York, NY, USA, 2004, pp. 315–321, ACM.

[3] P.R. White, W.B. Collis, S. Robinson, and A. Kokaram, "Inference matting," nov. - 1 dec. 2005, pp. 168 – 172.

[4] Anat Levin, Dani Lischinski, and Yair Weiss, "A closed form solution to natural image matting," in *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 61–68.

[5] A. Criminisi, A. Blake, C. Rother, J. Shotton, and P. H. Torr, "Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming," *Int. J. Comput. Vision*, vol. 71, no. 1, pp. 89–110, 2007.

[6] Jiejie Zhu, Miao Liao, Ruigang Yang, and Zhigeng Pan, "Joint depth and alpha matte optimization via fusion of stereo and time-of-flight sensor," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 0, pp. 453–460, 2009.

[7] Oliver Wang, Jonathan Finger, Qingxiong Yang, James Davis, and Ruigang Yang, "Automatic natural video matting with depth," in *PG '07: Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*, Washington, DC, USA, 2007, pp. 469–472, IEEE Computer Society.

[8] *Numerical Bayesian Methods Applied to Signal Processing*, SpringerVerlag, Springer Series in Statistics and Computing, 1996.

[9] Michael Bleyer and Margrit Gelautz, "Simple but effective tree structures for dynamic programming-based stereo matching," in *VISAPP (2)*, 2008, pp. 415–422.