AUTOMATIC TRIMAP EXTRACTION FOR EFFICIENT ALPHA MATTING BASED ON GRADIENT FIELD TRANSFORMS

Sang Min Yoon and Holger Graf
GRIS, TU Darmstadt, Rundeturmstrasse 10, Darmstadt, Germany
ZGDV, Computer Graphics Center, Rundeturmstrasse 10, Darmstadt, Germany

Keywords: Alpha matting, Trimap extraction, Gradient field.

Abstract: Image/Video Matting aims at solving the problem of accurate foreground estimation from a given background within still images or video sequences. The standard alpha matting method starts from a trimap, which separates an input image into three regions: definitely foreground, definitely background, and unknown regions. This paper presents an automatic trimap extraction based on an affine transformation of gradient fields in order to achieve an improved and robust Image/Video matting method. A gradient field based background and foreground segmentation technique provides a trimap extraction, which is robust to changing light conditions within semi-transparent objects. Our proposed background subtraction is based on affine transformed gradient projections of the input and background image and removes the background texture from a given image, preserving the texture of the foreground objects. The presented automatic trimap extraction method reduces the manual labor work in extracting and embedding target objects into a new background image or video sequences and might find its application within the broadcasting or movie industry.

1 INTRODUCTION

Image/Video matting, which deals with an extraction of foreground objects from a background image by a pixel with color or opacity segmentation, has been extensively studied during the last twenty years. Traditional matting techniques segment the foreground object of a given image or video sequences by its color and opacity characteristics. Current image/video matting technologies and systems try to efficiently extract a high degree of mattes from a still image or a video sequence and is mainly used in the film production and broadcasting systems. Here, image matting defines a given input image as the composite of a foreground layer and a background layer, and combines those by using a linear blending of opacity values in each pixel (Wang et al, 2007). It was first mathematically defined by (Smith et al, 1996) and models an observed image I as a convex combination of a foreground image F and a background image B by using the alpha matte α:

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

where α is a value within [0,1]. From this equation (1), the matting process has to solve an inverse problem with several unknowns, only given three constraints. The task of alpha matting is hence, to recover the value of α, B, F at every pixel. To properly extract meaningful foreground objects from equation (1), several matting approaches start to segment the input image into three regions which is referred to as a trimap: definitely foreground, definitely background, and unknown regions. If α = 1 or 0, we classify a pixel within an image as 'definitely foreground' or 'definitely background' respectively. One of the important factors affecting the performance of a matting algorithm is the accuracy of the trimap. The trimap reduces the dimension of the solution space for the matting problem, and leads the matting algorithm to generate user-desired results.

In this paper, we describe an automatic trimap extraction methodology by developing an affine transform of gradient fields between background image and a given input image for a robust alpha matting process. Our proposed background subtraction system is robust in extracting the foreground regions within illumination varying environments caused by changing light conditions. Figure 1 shows the flowchart for automatically extracting a trimap from a given image, based on background subtraction and a closed-form based alpha matting technique (Levin et al, 2006).

Yoon S. and Graf H. (2009).
AUTOMATIC TRIMAP EXTRACTION FOR EFFICIENT ALPHA MATTING BASED ON GRADIENT FIELD TRANSFORMS.
In *Proceedings of the Fourth International Conference on Computer Graphics Theory and Applications*, pages 52-57
DOI: 10.5220/0001766800520057

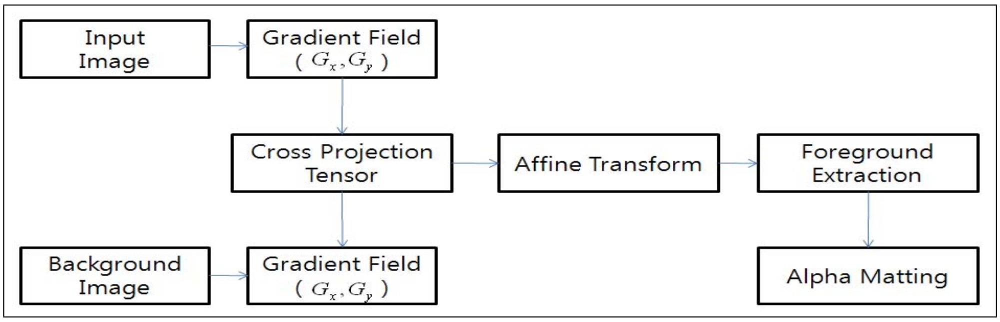


Figure 1: Flowchart of our proposed automatic trimap extraction for efficient alpha matting by transformed gradient fields using cross-projection tensors.

An affine transform of the cross projection tensors is computed in order to extract the foreground elements from a given input image automatically and separates it into the 'definitely foreground', the 'definitely background' and 'unknown regions'.

2 PREVIOUS RESEARCH

In this section, we briefly describe some state of the art research for image/video matting, automatic trimap extraction and efficient background subtraction methods.

2.1 Alpha Matting

An early parametric sampling based matting algorithm was proposed by (Ruzon et al, 2000), whose approach is based on a manifold, connecting the "frontiers" of each object's color distribution. Based on his approach, Bayesian matting (Chuang et al, 2001) uses a continuously sliding window for neighborhood definition, which marches inwards from the foreground and background regions. In order to build color distributions, the algorithm uses foreground and background samples additionally to computed F, B and α values. Thus, every pixel within a neighborhood region will contribute to model the foreground and background Gaussian. However, the parametric sampling based matting is weak when the background color is non-Gaussian. The "knockout matting" method (Berman et al, 2000) has been developed to avoid the disadvantages of parametric sampling matting by a weighted average of known foreground and background pixels. Poisson matting (Sun et al, 2004) solves a Poisson equation for the matte by assuming that the foreground and background are slowly varying compared to the matte. This algorithm interacts closely with the user by beginning from a hand-painted trimap offering painting tools to correct errors within the matte. Defocus matting (McGuire et al, 2005) works with pulling the mattes automatically from video sequences captured with co-axial cameras within different depths of field and plane of focus. Instead of requiring a carefully specified trimap, some recently proposed matting approaches allow the user to specify a few foreground and background scribbles as user input to extract a matte. However, in order to reduce user involvement, automatic trimap extraction, is one of the most important issues in image/video matting. Optical flow based background subtraction (Chuang et al, 2002) or depth map extraction (Joshi et al, 2007) is suggested to separate foreground, and background automatically. A recently published spectral matting algorithm (Levin et al, 2007) can automatically extract a matte from an input image without any user input and provides an efficient video matting video matting for even complex scenes, using an optical flow to separate foreground regions from background image.

2.2 Background Subtraction

The principle of a background subtraction methodology is to detect moving objects by building the difference between the current frame and a reference frame. A comprehensive overview and in depth literature review on background subtraction techniques can be found in (Piccardi et al, 2004). Several methods for performing background subtraction try to effectively estimate the background model from temporally trained sequences of images. (Wren et al, 1997) has proposed to model the background independently at each pixel which is based on a Gaussian probability density function. (Stauffer et al, 1999) extended the uni-modal background subtraction approach by using an adaptive multi-modal background subtraction method that modelled the pixel color as a mixture of Gaussians. (Oliver et al, 2000) used an eigen space model for background subtraction. Background subtraction techniques which combine multiple cues such as color and depth maps are also used for video surveillance and monitoring system (Barotti et al, 2003).

3 AUTOMATIC TRIMAP EXTRACTION

Traditional background subtraction algorithms (Wren et al, 1997, Oliver et al, 2000, Elgammal et al, 2000, Barotti et al, 2003, Han, 2004) which are based on frame differences cannot extract the illumination and reflectance variable foreground objects and fail to extract the foreground objects that have similar color values to the background. However, an image gradient field based on projection tensors provides a way of removing scene texture edges from images within the illumination parameter space. We now show how to remove the background from a given image by transforming its gradient field using cross projection tensors obtained from a background image of the same scene. The final trimap for alpha matting is obtained by a 2D integration of the modified gradient field from each color channel.

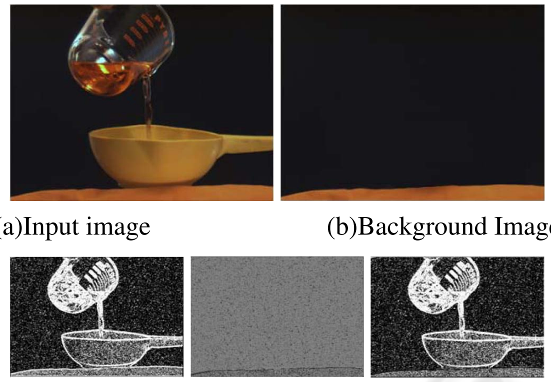
3.1 Background Subtraction from Transformed Gradient Fields

To extract the foreground regions from a given image, our proposed background subtraction in the gradient fields is motivated by (Agrawal et al, 2006) whose method removes the scene texture which is corresponding to the background. From a given image I the smoothed structure tensor, G_σ of the gradient image ∇I is defined as the convolution of the gradient field and Gaussian kernel. The smoothed structure tensor with Gaussian kernel reduces the local noise and aliasing. Equation (2) shows the structured tensor from the gradient field and Gaussian kernel.

$$G_\sigma = (\nabla I \nabla I^T) * K_\sigma = \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix} * K_\sigma \tag{2}$$

where * denotes the convolution and K_σ is a normalized 2D Gaussian kernel of variance σ. Equation (2) can be decomposed as eigenvectors and eigenvalues of the input and background image as shown in equation (3) in order to extract the local intensity structure within the image (Aubert et al, 2002).

$$G_\sigma = [v_1 v_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix} \tag{3}$$



(a)Input image (b)Background Image

(c) D^{11} of D (d) D^{12} of D (e) D^{22} of D

Figure 2: (a) is a given input image for alpha matting (b) a background image. (c), (d), and (e) are components of a diffusion tensor to extract a foreground object.

where v_1, v_2 denote the eigenvectors corresponding to the eigenvalues λ_1, λ_2 respectively and $\lambda_1 \leq \lambda_2$. The eigenvalues and eigenvectors of G_σ provide information about the local intensity structures within the image.

$$D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} = [u_1 u_2] \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \end{bmatrix} \tag{4}$$

The field of diffusion tensors at each pixel can be described through a 2×2 symmetric, positive definite matrix (Weickert, 2004). The diffusion tensors D by selecting its eigenvectors u_1, u_2 and eigenvalues μ_1, μ_2 based on eigenvalues and eigenvectors of G_σ. D is then obtained as equation (4). Based on the cross projection tensor of the gradient field between the input and background image, we suppress the texture information of the background, still preserving the texture of the background. With extracting the transformed gradient field using cross projection tensors, we integrate the transformed gradient field of each channel of the image.

From the background of gradient fields and diffusion function, we will focus on gradient projection of a given image, I_o, and background image,I_b. By transforming ∇I_o with cross projection tensor D^{OB}, which is the diffusion tensor between ∇I_o and ∇I_b, we can remove all edges from I_o which are in I_b and retain all edges in I_o which are not in I_b. Thus, scene texture edge from an input image can be removed by transforming its gradient field using cross projection tensors obtained from a background image of the same scene. If we define an input image, I_o and an background image, I_b, the smoothed structure tensors of each image are denoted as G_σ^o and G_σ^b, respectively.

Figure 2-(a) and (b) show the input image and the background image. Figure 2-(c),(d), and (e) show the

extracted D_{11}, $D_{12} = D_{21}$, and D_{22} of the cross projection tensors.

3.2 Trimap Extraction from Background Subtraction

From the cross projected tensors for each color channel, we extract the trimap of the input image for an efficient alpha matting. Extracted foreground regions based on gradient fields include definitely foreground and unknown region by illumination or intricately shape of the target object in the input image is the difference of the gradient field between the input image and the affine transformed gradient field by tensor projection. The definitely foreground region based on the roughly extracted foreground is separated into the adaptive threshold that represents the channel difference and gradient difference. Figure 3 shows the extracted trimap from the input image in the case of fluid and fire image. Within the trimaps which are shown in figure 3, we assigned to definitely background element as 0, definitely foreground elements are assigned to 255, and unknown regions to 128. As shown in figure 3, foreground regions which are extracted from the background subtraction are separated as definitely foreground and unknown regions.

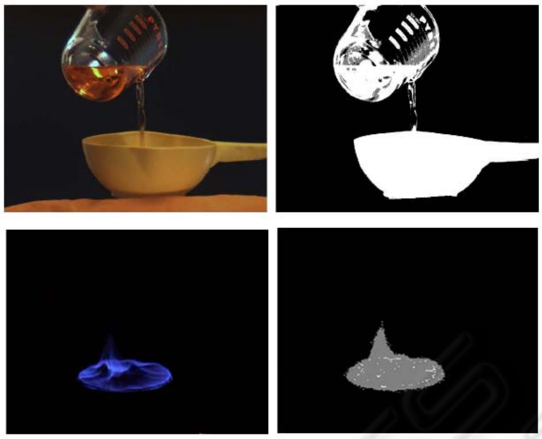


Figure 3: Extracted trimap from efficient background subtraction based on gradient fields in the case of fluid and fire.

4 CLOSED-FORM MATTING

From the various image/video matting algorithms from an image or video sequences, we tested a closed-form based alpha matting with our automatically extracted trimap. The closed-form matting algorithm which was recently presented by (Levin et al., 2006) is based on explicitly deriving a cost function from local smoothness assumptions on foreground and background. It eliminates F and B, yielding a quadratic cost function in alpha, which can be easily solved as a sparse linear system of equations. The mathematical alpha matting equation which is represented in equation (1) can be re-written if each F and B is a linear mixture of two colors over a small window around each pixel as:

$$\alpha = \sum_c a^c I_i^c + b, \forall i \in w \tag{5}$$

where equation (5) is referred to as linear color channels, and a^c and b are constants within the window w. The matting cost function J is only dependent on α as shown in equation (6).

$$J(\alpha, a, b) = \sum_{j \in I} \Big(\sum_{i \in w_j} (\alpha_i - \sum_c a_j^c I_i^c - b_j)^2 + \varepsilon \sum_c a(c^2)_j \Big) \tag{6}$$

From the equation (5), a^c and b can be eliminated from the cost function J, yielding a quadratic cost in the α alone:

$$J(\alpha) = \alpha^T L \alpha \tag{7}$$

here L is an $N \times N$ operator, whose (i, j)-th element denoted to:

$$\sum_{k|(i,j) \in w_k} \Big(\delta_(ij) - \frac{1}{|w_k|}(1 + (I_i - \mu_k)(\Sigma_k + \frac{\varepsilon}{|w_k|} I_3)^{-1})(I_i - \mu_k)\Big) \tag{8}$$

where Σ_k is a 3×3 covariance matrix, μ_k is a 3×1 mean vector of the colors in window w_k, and I^3 is the 3×3 identity matrix. The operator L, which is called the Matting Laplacian, is the most important analytic result from this approach. The affinity defined in Equation (6) and the one defined in Equation (8) share the same property that nearby pixels with similar colors have high affinity values, while nearby pixels with different colors have small affinity values. Figure 4 shows the recovered alpha values and embedding into a new background. As shown in figure 4, the illuminated area within the original target object is represented in a new background and the semi-transparent object, e.g. fire, that contains many holes are also robustly embedded into a new background image.

5 EXPERIMENTS

We have implemented the alpha matting algorithm from the gradient fields and conducted some experiments on a standard PC with Pentium4 1.2GHz in order to show the efficiency and robustness for our

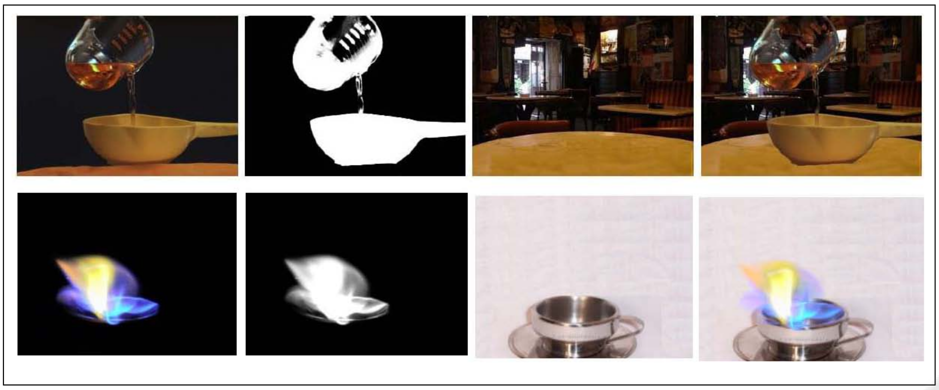


Figure 4: Alpha extraction and matting in a new background.

Figure 5: Comparison of alpha matting between our proposed method and previous matting with depth map(McGuire et al, 2006). Left image is the original input image, the center image is the foreground.

automatic trimap extraction and alpha matting by comparing it to previous research of background subtraction, trimap extraction, and alpha matting method. Our first benchmark computes the background subtraction method based on a kernel density estimation (Han, 2004). Based on the original image in figure 5 (left), we show the results of our work within the gradient field domain (center) and the extracted foregrond objects by (Han, 2004)(right). The light reflected area within the image is well reconstructed in our approach compared to the other.

We lead the alpha matting experiment from an automatically extracted trimap. Especially trimap extraction of the semi-transparent object such as fire or fluids is difficult because there are many holes within the interior of the target object. Figure 6 shows the alpha matting of our proposed method and a previous alpha matting method based on dual cameras (McGuire, 2006).

6 CONCLUSIONS AND FUTURE WORKS

Within this paper, we have presented an automatic trimap extraction using a new form of background subtraction which is robust within illumination variant environments. The extracted trimap is applied to a closed-form based alpha matting process as we would like to show, that an automatic trimap extraction within a gradient field is more efficient than other background subtraction methods which are based on building frame differences or applying optical flow techniques to semi-transparent objects in view of changing illumination and/or light reflectance. Our experiments proved that the proposed method is very robust in alpha matting of semi-transparent objects. Physical phenoma, such as fire or fluid, have been chosen as sample data due to their nature in time varying shape changes, comprising large wholes and unknown regions within their inner shape, and lighting characteristics. We are confident that this approach could reduce the labor work in extracting and embedding target objects into a new background image or video sequences.

However, a closed-form based alpha matting process which follows our automatically extracted trimap extraction still takes too much time when the unknown regions in a given image are large. We will thus transfer our proposed algorithm into a GPU environment in order to reduce the processing time by parallel calculation of the gradient fields and small neighborhood window processing of unknown regions for an efficient and fast alpha extraction.

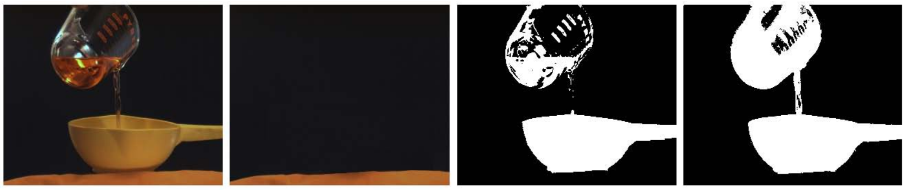


Figure 6: Comparison of background subtraction between gradient field based affine transform and kernel density estimation. Gradient field based background subtraction is more efficient in the area of illumination change or light's semi-transmission.

ACKNOWLEDGEMENTS

We have experiments with the fire image data from the Informatik, Max-Plank Institute (Ihrke et al, 2005) and fluid image data of the department of computer science at Williams College (McGuire et al, 2006).

REFERENCES

Agrawal, A., Raskar, R., Chellappa, R., 2006. Edge suppression by gradient field transformation using cross-projection tensors. *In proceeding of IEEE CVPR*

Aubert, G., Kornprobst, G., 2002. Mathematical Problems in Image Processing: Partial Differential Equations and the Calcuus of Varations. *Applied Mathematical Series, Springer-Verlag.*

Barotti, S., Lombardi, L., Lombardi, P., 2003. Multi-Module Switching and Fusion for Robust Video Surveillance. *In proceeding of IEEE International Conference on Image Analysis and Processing.*

Berman, A., Vlahos, P., Dadourian, A., 2000. Comprehensive method for removing from an image the background surrounding. *U.S. Patent.*

Chuang, Y.-Y., Curless, B., Salesin, D. H., Szeliski, R., 2001. A bayesian approach to digital matting. *In proceeding of IEEE CVPR.*

Chuang, Y., Agarwala, A., Curless, B., Szeliski, R., 2002. Video matting of complex scenes. *ACM Transaction Graph.*

Elgammal, A., Harwood, D., Davis, L., 2000. Non-parametric model for background subtraction. *In proceeding of ECCV.*

Han, B., Comaniciu, D., Davis, L., 2004. Sequential kernel density approximation through mode propagation: applications to background modeling. *In proceeding of ACCV.*

Ihrke, I., Magnor, M., 2005. GrOVis-Fire: A Multi-Video Sequence for Volumetric Reconstruction and Rendering Research. *http://www.mpi-sb.mpg.de/~ihrke/Projects/Fire/GrOVis_Fire/*

Joshi, N., Matusik, W., Freeman, W. F., 2007. Exploring Defocus Matting: Non-Parametric Acceleration, Super-Resolution, and Off-Center Matting . *IEEE Computer Graphics and Applications.*

Levin, A., Linschinski, D., Weiss, Y., 2006. A closed form solution to natural image matting. *In proceeding of IEEE CVPR.*

Levin, A., Rav-Acha, A., Lischinski, D., 2007. Spectral Matting. *In proceeding of IEEE CVPR.*

McGuire, M., Matusik, W., Yerazunis, W., 2006. Practical, Real-time Studio Matting using Dual Imagers. *In European Symposium on Rendering.*

McGuire, M., Matusik, W., Pfister, H., Hughes, J, F., Durand, F., 2005. Defocus Video Matting. *In proceeding of ACM SIGGRAPH.*

land, A. P., 2000. A bayesian computer vision system for modeling human interactions. *IEEE Trans. on PAMI.*

Piccardi, M., 2004. Background subtraction techniques: a review. *IEEE International Conference on System, Man, and Cybernetics.*

Ruzon, M., Tomasi, C., 2000. Alpha estimation in natural images. *In proceeding of IEEE CVPR.*

Smith, A., Blinn, J., 1996. Blue screen matting. *In proceeding of ACM SIGGRAPH.*

Stauffer, C., Grimson, W., 1999. Adaptive background mixture models for real-time tracking. *In proceeding of IEEE CVPR.*

Sun, J., Jia, J., Tang, C.-K., Shum, H.-Y., 2004. Poisson matting. *In proceeding of ACM SIGGRAPH.*

Tschumperle, D., 2002. PDE's based Regularization of Multivalued Images and Applications. *PhD Thesis, Universite de Nice-Sophia Antipolis.*

Wang, J., Cohen, M. F., 2007. Image and video matting: a survey. *Foundations and Trends in Computer Graphics and Vision.*

Weickert, J., 1997. A review of nonlinear diffusion filtering. *In Scale-Space Theory in Computer Vision, Springer.*

Wren, C., Azarbayejani, A., Darrell, T., Pentland, A. P., 1997. Pfinder: real-time tracking of the human body. *IEEE Trans on PAMI.*

Zomet, A., Peleg, S., 2002. Multi-sensor super resolution. *IEEE workshop on Application of Computer Vision.*