# Image and Video Matting: A Survey

# Image and Video Matting: A Survey

---

**Jue Wang**

*Adobe Systems Incorporated*
*801 North 34th Street*
*Seattle, WA 98103*
*USA*

*juewang@adobe.com*

**Michael F. Cohen**

*Microsoft Research*
*One Microsoft Way*
*Redmond, WA 98052*
*USA*

*michael.cohen@microsoft.com*

now
the essence of knowledge

Boston – Delft

Full text available at: http://dx.doi.org/10.1561/0600000019

# Foundations and Trends® in Computer Graphics and Vision

*Published, sold and distributed by:*
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

*Outside North America:*
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is J. Wang and M. F. Cohen, Image and Video Matting: A Survey, Foundations and Trends® in Computer Graphics and Vision, vol 3, no 2, pp 97–180, 2007

ISBN: 978-1-60198-134-9
© 2008 J. Wang and M. F. Cohen

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

**Foundations and Trends® in**
**Computer Graphics and Vision**
Volume 3 Issue 2, 2007

# Editorial Scope

**Foundations and Trends® in Computer Graphics and Vision** will publish survey and tutorial articles in the following topics:

- Rendering: Lighting models; Forward rendering; Inverse rendering; Image-based rendering; Non-photorealistic rendering; Graphics hardware; Visibility computation
- Shape: Surface reconstruction; Range imaging; Geometric modelling; Parameterization;
- Mesh simplification
- Animation: Motion capture and processing; Physics-based modelling; Character animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape Representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and Video Retrieval
- Video analysis and event recognition
- Medical Image Analysis
- Robot Localization and Navigation

**Information for Librarians**

Foundations and Trends® in Computer Graphics and Vision, 2007, Volume 3, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Foundations and Trends® in
Computer Graphics and Vision
Vol. 3, No. 2 (2007) 97–180

DOI: 10.1561/0600000019

# Image and Video Matting: A Survey

**Jue Wang**[1] **and Michael F. Cohen**[2]

[1] *Adobe Systems Incorporated, 801 North 34th Street, Seattle, WA 98103, USA, juewang@adobe.com*
[2] *Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA, michael.cohen@microsoft.com*

## Abstract

Matting refers to the problem of accurate foreground estimation in images and video. It is one of the key techniques in many image editing and film production applications, thus has been extensively studied in the literature. With the recent advances of digital cameras, using matting techniques to create novel composites or facilitate other editing tasks has gained increasing interest from both professionals as well as consumers. Consequently, various matting techniques and systems have been proposed to try to efficiently extract high quality mattes from both still images and video sequences.

This survey provides a comprehensive review of existing image and video matting algorithms and systems, with an emphasis on the advanced techniques that have been recently proposed. The first part of the survey is focused on image matting. The fundamental techniques shared by many image matting algorithms, such as color sampling methods and matting affinities, are first analyzed. Image matting techniques are then classified into three categories based on their underlying methodologies, and an objective evaluation is conducted to reveal the advantages and disadvantages of each category. A unique Accuracy vs. Cost analysis is presented as a practical guidance for readers to properly choose matting tools that best fit their specific requirements and constraints.

The second part of the survey is focused on video matting. The difficulties and challenges of video matting are first analyzed, and various ways of combining matting algorithms with other video processing techniques for building efficient video matting systems are reviewed. Key contributions, advantages as well as limitations of important systems are summarized.

Finally, special matting systems that rely on capturing additional foreground/background information to automate the matting process are discussed. A few interesting directions for future matting research are presented in the conclusion.

# Contents

# 1

# Introduction

## 1.1 The Matting Problem

Extracting foreground objects from still images or video sequences plays an important role in many image and video editing applications, thus it has been extensively studied for more than 20 years. Accurately separating a foreground object from the background involves determining both full and partial pixel coverage, also known as *pulling a matte*, or *foreground matting*. This problem was mathematically established by Porter and Duff in 1984 [29]. They introduced the alpha channel as the means to control the linear interpolation of foreground and background colors for anti-aliasing purposes when rendering a foreground over an arbitrary background. Mathematically, the observed image $I_z$ ($z = (x,y)$) is modeled as a convex combination of a foreground image $F_z$ and a background image $B_z$ by using the alpha matte $\alpha_z$:

$$I_z = \alpha_z F_z + (1 - \alpha_z) B_z, \tag{1.1}$$

where $\alpha_z$ can be any value in [0,1]. If $\alpha_z = 1$ or 0, we call pixel $z$ *definite foreground* or *definite background*, respectively. Otherwise we call pixel $z$ *mixed*. In most natural images, although the majority of pixels are either definite foreground or definite background, accurately estimating alpha values for mixed pixels is essential for fully separating the foreground from the background.

Given only a single input image, all three values $\alpha$, $F$, and $B$ are unknown and need to be determined at every pixel location. The known information we have for a pixel are the three dimensional color vector $I_z$ (assuming it is represented in some 3D color space), and the unknown variables are the three dimensional color vectors $F_z$ and $B_z$, and the scalar alpha value $\alpha_z$. Matting is thus inherently an under-constrained problem, since 7 unknown variables need to be estimated from 3 known values. Most matting approaches rely on user guidance and prior assumptions on image statistics to constrain the problem to obtain good estimates of the unknown variables. Once estimated correctly, the foreground can be seamlessly composed onto a new background, by simply replacing the original background $B$ with a new background image $B'$ in Equation (1.1).

## 1.2 Binary Segmentation vs. Matting

If we constrain the alpha values to be only 0 or 1 in Equation (1.1), the matting problem then degrades to another classic problem: binary image/video segmentation, where each pixel fully belongs to either foreground or background. This problem has been extensively studied since early 1960s, resulting in a large volume of related literature. Although matting is modeled as a more general problem than binary segmentation, which is theoretically harder to solve, most existing matting algorithms avoid the segmentation problem by having a *trimap* as another input in addition to the original image. The trimap may be manually specified by the user, or produced by other binary segmentation approaches. The trimap reduces the dimension of the solution space of the matting problem, and leads the matting algorithms to generate user-desired results.

Although binary segmentation and alpha matting are closely coupled problems, in this survey for image matting we will assume that a rough foreground segmentation is given, thus we mainly focus on how to accurately estimate alpha values for truly mixed pixels. We will however discuss binary segmentation techniques in the context of video matting since they play a more central role in recent video matting systems.

## 1.3 The Trimap

Without any additional constraints, it is obvious that the total number of valid solutions to Equation (1.1) is infinite. For a trivial solution, one can set all $\alpha_z$s to be 1 and all $F_z$s to be identical to $I_z$s, which simply means the whole image is fully occupied by the foreground. Of course this solution is probably not consistent with what a human being will perceive from the input image. To properly extract semantically meaningful foreground objects, almost all matting approaches start by having the user segment the input image into three regions: definitely foreground $R_f$, definitely background $R_b$, and unknown $R_u$. This three-level pixel map is often referred to as a *trimap*. The matting problem is thus reduced to estimating $F$, $B$, and $\alpha$ for pixels in the unknown region based on known foreground and background regions. An example of a trimap is shown in Figure 1.1.

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. This intrinsically defines a very coarse trimap by marking the majority pixels (pixels have not been touched by the user) as unknowns.

One of the important factors effecting the performance of a matting algorithm is how accurate the trimap is. Ideally, the unknown region in the trimap should only cover truly mixed pixels. In other words, the unknown region around the foreground boundary should be as thin

Fig. 1.1 A matting example. From left to right: input image; user specified trimap; extracted matte; estimated foreground colors; a new composite. Results are generated by the Robust Matting algorithm [49].

as possible to achieve the best possible matting results. This is somewhat obvious since the more accurate the trimap is, the less number of unknown variables need to be estimated, and the more known foreground and background information is available to use. However, accurately specifying a trimap requires significant amounts of user effort and is often undesirable in practice, especially for objects with large semi-transparent regions or holes. Thus a big challenge for designing a successful matting algorithm is how to achieve a good trade-off between the accuracy of the matte and the amount of the user effort required. As we will see later, different algorithms have totally different characteristics in this accuracy–efficiency space.

It is worth mentioning that the recently proposed Spectral matting algorithm [22] can automatically extract a matte from an input image without any user input. However, as the authors agreed, the automatic approach has a number of limitations including erroneous results for images with highly textured backgrounds. Thus in practice, user specified trimaps are typically necessary to achieve high quality matting results.

## 1.4 The User Interface

A properly designed user interface is critical to the success of an interactive system. Surprisingly, although the matting problem has been studied for more than two decades, very little research has been done on exploring good user interfaces for the matting task. Most of the existing matting systems work in an offline mode, where in the interactive loop, the user first specifies a trimap, that invokes matting algorithms to compute a matte. If the result is not satisfactory, the user then refines the trimap and runs the algorithm again. On the other hand, recently proposed matting algorithms mainly focus on how to improve the quality of the matte by introducing more sophisticated analysis and optimization methods, thus they are generally slow. As a result, the interactive loop described above can be very time-consuming and inefficient.

The recently proposed Soft Scissors system [46] demonstrates the possibility of a realtime matting user interface. In this system, a trimap is created incrementally by the user with the aid of a polarized brush stroke (stroke with foreground/background boundary conditions) with dynamically updated parameters. Alpha values of pixels inside the brush stroke are computed in realtime as the user paints along the foreground edge. The instant feedback allows the user to immediately see what the foreground will look like over a new background. This approach opens many new possibilities for creating more efficient and intelligent matting user interfaces.

Another interesting image matting interface is the "components picking" interface proposed in [22]. In this approach a set of fundamental fuzzy matting components are automatically extracted from an input image, based on analyzing the smallest eigenvectors of a suitably defined Laplacian matrix. The user then selects proper components to form the foreground object using simply a few mouse clicks. However, in the case that the automatically computed components are not accurate enough, how to fine adjust the resulting matte on pixel level is unknown. One can imagine combining this approach with other matting interfaces for generating more accurate results.

Designing efficient user interfaces for video matting is certainly a more challenging task. Existing video matting interfaces can be classified into two categories: keyframe-based and volume-based approaches. Systems in the first category allow users to provide inputs on manually or automatically selected keyframes which are sparsely distributed in the input sequence, then try to automatically propagate them into intermediate frames to create a full set of constraints. Volume-based systems treat the video data as a 3D spatio-temporal video cube and allow users to directly marking pixels on extruded surfaces from the 3D cube. Details of these systems will be discussed in Section 6.

## 1.5 Matting with Extra Information

In early matting systems, the input image is often captured against a single or multiple constant-colored background(s), known as *blue screen matting*. As shown in these approaches, knowing the background greatly reduces the difficulty for extracting an accurate matte.

For better matting results on natural images and video, special imaging systems have been designed to provide additional information or constraints to matting algorithms, such as using flash or non-flash image pairs [39], camera arrays [20], and multiple synchronized video streams [26]. Leveraging these additional sources of information, lower complexity matting algorithms can be designed to achieve fast and accurate matting. These approaches will be discussed in detail in Section 7.

# References

[1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," in *Proceedings of ACM SIGGRAPH*, pp. 294–302, 2004.

[2] A. Agarwala, A. Hertzmann, D. H. Salesin, and S. M. Seitz, "Keyframe-based tracking for rotoscoping and animation," in *Proceedings of ACM SIGGRAPH*, pp. 584–591, 2004.

[3] X. Bai and G. Sapiro, "A geodesic framework for fast interactive image and video segmentation and matting," in *Proceedings of IEEE ICCV*, 2007.

[4] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[5] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[6] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient n-d image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[7] W. L. Briggs, V. E. Henson, and S. F. Mccormick, *A Multigrid Tutorial*. Philadelphia: Society for Industrial and Applied Mathematics, 2000.

[8] Y.-Y. Chuang, A. Agarwala, B. Curless, D. Salesin, and R. Szeliski, "Video matting," in *Proceedings of ACM SIGGRAPH*, pp. 243–248, 2002.

[9] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski, "A bayesian approach to digital matting," in *Proceedings of IEEE CVPR*, pp. 264–271, 2001.

[10] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[11] C. CORPORATION, "Knockout user guide," 2002.

[12] J. Dennis, J. Robert, and B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, 1983.

[13] I. S. Duff, A. M. Erisman, and J. K. Reid, *Direct Methods for Sparse Matrices*. Oxford: Clarendon Press, 1986.

[14] V. Eijkhout, "Overview of iterative linear system solver packages," Technical Report, USA, Knoxville, TN: University of Tennessee, 1998.

[15] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," in *Proceedings of ACM SIGGRAPH*, pp. 673–678, 2004.

[16] L. Grady, T. Schiwietz, S. Aharon, and R. Westermann, "Random walks for interactive alpha-matting," in *Proceedings of VIIP 2005*, pp. 423–429, 2005.

[17] Y. Guan, W. Chen, X. Liang, Z. Ding, and Q. Peng, "Easy matting: A stroke based approach for continuous image matting," in *Proceeding of Eurographics*, pp. 567–576, 2006.

[18] X. He and P. Niyogi, "Locality preserving projections," in *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, 2003.

[19] J. Jia, J. Sun, C.-K. Tang, and H.-Y. Shum, "Drag-and-drop pasting," in *Proceedings of ACM SIGGRAPH*, 2006.

[20] N. Joshi, W. Matusik, and S. Avidan, "Natural video matting using camera arrays," in *Proceedings of ACM SIGGRAPH*, pp. 779–786, 2006.

[21] A. Levin, D. Lischinski, and Y. Weiss, "A closed form solution to natural image matting," in *Proceedings of IEEE CVPR*, pp. 228–242, 2006.

[22] A. Levin, A. Rav-Acha, and D. Lischinski, "Spectral matting," in *Proceedings of IEEE CVPR*, 2007.

[23] A. Levin, A. Rav-Acha, and D. Lischinski, "Spectral matting," Hebrew University Technical Report, 2007.

[24] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," in *Proceedings of ACM SIGGRAPH*, pp. 303–308, 2004.

[25] Y. S. Y. Li and H. Y. Shum, "Video object cut and paste," in *Proceedings of ACM SIGGRAPH*, pp. 595–600, 2005.

[26] M. McGuire, W. Matusik, H. Pfister, J. F. Hughes, and F. Durand, "Defocus video matting," in *Proceedings of ACM SIGGRAPH*, pp. 567–576, 2005.

[27] Y. Mishima, "Soft edge chroma-key generation based upon hexoctahedral color space," U.S. Patent 5,355,174, 1993.

[28] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, and M. Cohen, "Digital photography with flash and no-flash image pairs," in *Proceedings of ACM SIGGRAPH*, pp. 664–672, 2004.

[29] T. Porter and T. Duff, "Compositing digital images," in *Proceedings of ACM SIGGRAPH*, pp. 253–259, July 1984.

[30] R. Raskar, K. Tan, R. Feris, J. Yu, and M. Turk, "Non-photorealistic camera: Depth edge detection and stylized rendering using multi-flash imaging," in *Proceedings of ACM SIGGRAPH*, pp. 673–678, 2004.

[31] C. Rhemann, C. Rother, A. Rav-Acha, and T. Sharp, "High resolution matting via interactive trimap segmentation," in *Proceedings of IEEE CVPR*, 2008.

[32] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut — Interactive foreground extraction using iterated graph cut," in *Proceedings of ACM SIGGRAPH*, pp. 309–314, 2004.

[33] C. Rother, V. Kolmogorov, and A. Blake, "Image blending," Tech. Rep., European Patent EP1748389, 2007.

[34] M. A. Ruzon and C. Tomasi, "Alpha estimation in natural images," in *Proceedings of IEEE CVPR*, pp. 18–25, 2000.

[35] Y. Saad, *Iterative methods for sparse linear systems*. SIAM, 2003.

[36] H. Y. Shum, J. Sun, S. Yamazaki, Y. Li, and C. Tang, "Pop-up light field: An interactive image-based modeling and rendering system," in *ACM Transactions on Graphics*, vol. 23, pp. 143–162, 2004.

[37] A. R. Smith and J. F. Blinn, "Blue screen matting," in *Proceedings of ACM SIGGRAPH*, pp. 259–268, 1996.

[38] J. Sun, J. Jia, C.-K. Tang, and H.-Y. Shum, "Poisson matting," in *Proceedings of ACM SIGGRAPH*, pp. 315–321, 2004.

[39] J. Sun, Y. Li, S.-B. Kang, and H.-Y. Shum, "Flash matting," in *Proceedings of ACM SIGGRAPH*, pp. 772–778, 2006.

[40] R. Szeliski, "Fast surface interpolation using hierarchical basis functions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 6, pp. 513–528, 1990.

[41] R. Szeliski, "Locally adapted hierarchical basis preconditioning," in *Proceedings of ACM SIGGRAPH*, pp. 1135–1143, 2006.

[42] R. Szeliski and H. Y. Shum, "Creating full view panoramic mosaics and environment maps," in *Proceedings of ACM SIGGRAPH*, pp. 251–258, 1997.

[43] J. K. Udupa and S. Samarasekera, "Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation," *Graphical Models and Image Processing*, vol. 58, pp. 246–261, 1996.

[44] P. Villegas and X. Marichal, "Perceptually-weighted evaluation criteria for segmentation masks in video sequences," in *IEEE Transactions on Image Processing*, vol. 13, pp. 1092–1103, 2004.

[45] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 583–598, 1991.

[46] J. Wang, M. Agrawala, and M. Cohen, "Soft scissors: An interactive tool for realtime high quality matting," in *Proceedings of ACM SIGGRAPH*, 2007.

[47] J. Wang, P. Bhat, A. Colburn, M. Agrawala, and M. Cohen, "Interactive video cutout," in *Proceedings of ACM SIGGRAPH*, 2005.

[48] J. Wang and M. Cohen, "An iterative optimization approach for unified image segmentation and matting," in *Proceedings of ICCV 2005*, pp. 936–943, 2005.

[49] J. Wang and M. Cohen, "Optimized color sampling for robust matting," in *Proceedings of IEEE CVPR*, 2007.

[50] J. Wang and M. Cohen, "Simultaneous matting and compositing," in *Proceedings of IEEE CVPR*, 2007.

[51] Y. Weiss and W. T. Freeman, “On the optimality of solutions of the max-product belief propagation algorithm in arebitrary graphs,” *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 303–308, 2001.

[52] G. Wyszecki and W. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae.* New York, NY: John Wiley and Sons, 1982.

[53] C. Yang, R. Duraiswami, N. Gumerov, and L. Davis, “Improved fast gauss transform and efficient kernel density estimation,” in *Proceedings of IEEE ICCV*, 2003.

[54] Y. Zheng, C. Kambhamettu, J. Yu, T. Bauer, and K. Steiner, “Fuzzymatte: A computationally efficient scheme for interactive matting,” in *Proceedings of IEEE CVPR*, 2008.