Supplementary material

A Perceptually Motivated Online Benchmark for Image Matting

Christoph Rhemann, Carsten Rother, Jue Wang, Margrit Gelautz, Pushmeet Kohli, Pamela Rott

CVPR 2009

The following slides show further results which are not included in the paper.

Important: Note that the supplementary material (PDF slides) are not necessary to understand the paper.

To reduce the size of the document, images were slightly compressed. Thus small artefacts may be visible.

Please view the slides in full screen mode.

The following slides introduce the Ground Truth (GT) Database

Ground Truth Database Overview

Thumbnail size (~0.1MPixel)

Test dataset

Training dataset

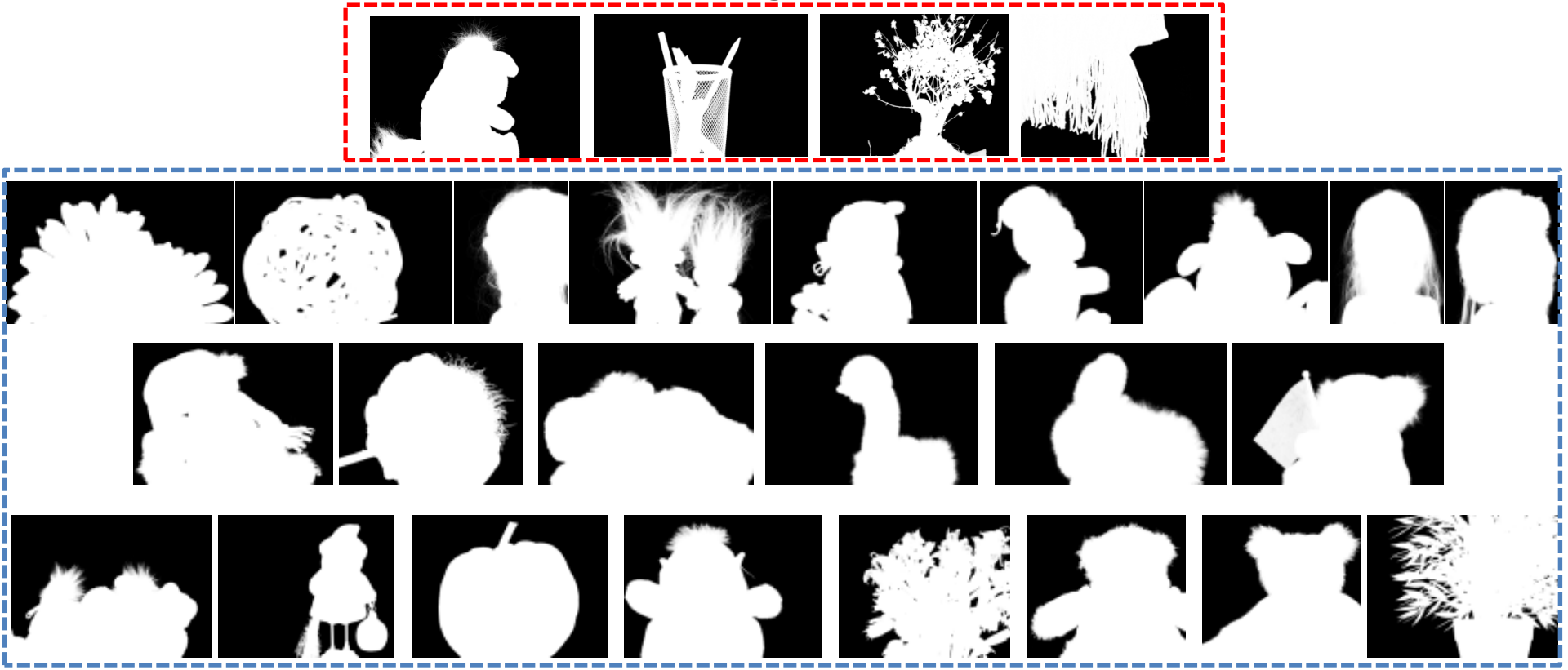
Composites from our dataset (marked in red) and from the database of [16] (marked in blue).
Images are downscaled – High resolution images are available at www.alphamatting.com

Ground Truth Database Overview

Thumbnail size (~0.1MPixel)

Test dataset

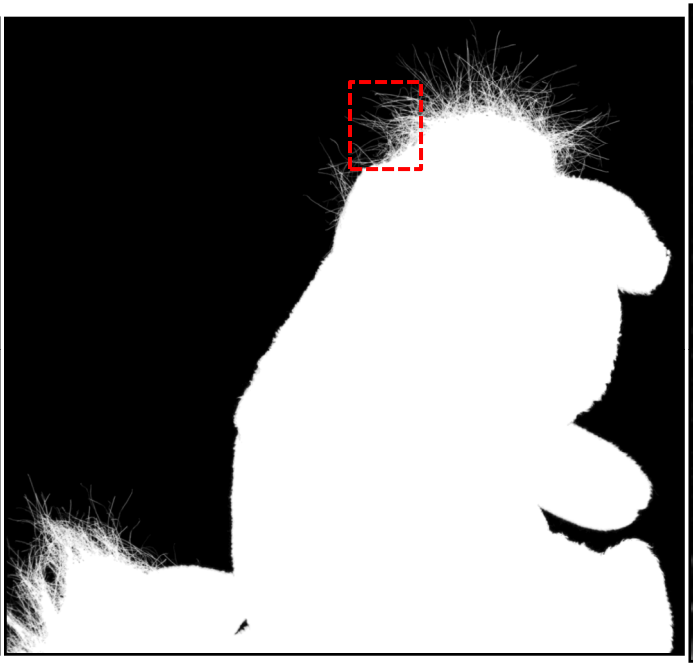
Training dataset



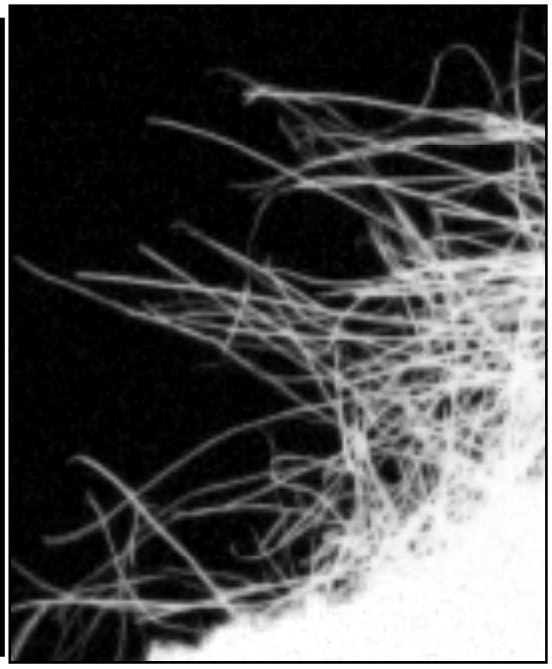
GT alpha from our dataset (marked in red) and from the database of [16] (marked in blue).
Images are downscaled – High resolution images are available at www.alphamatting.com

GT Database Close Up (1)

(a) Composite



(b) GT alpha



(c) Zoom-in of marked area in (b) (Full resolution)

Example from our new dataset

GT Database Close Up (2)

(a) Composite

(b) GT alpha

(c) Zoom-in of marked area in (b)

(Full resolution)

Example image from [16]

Qualitative Comparison

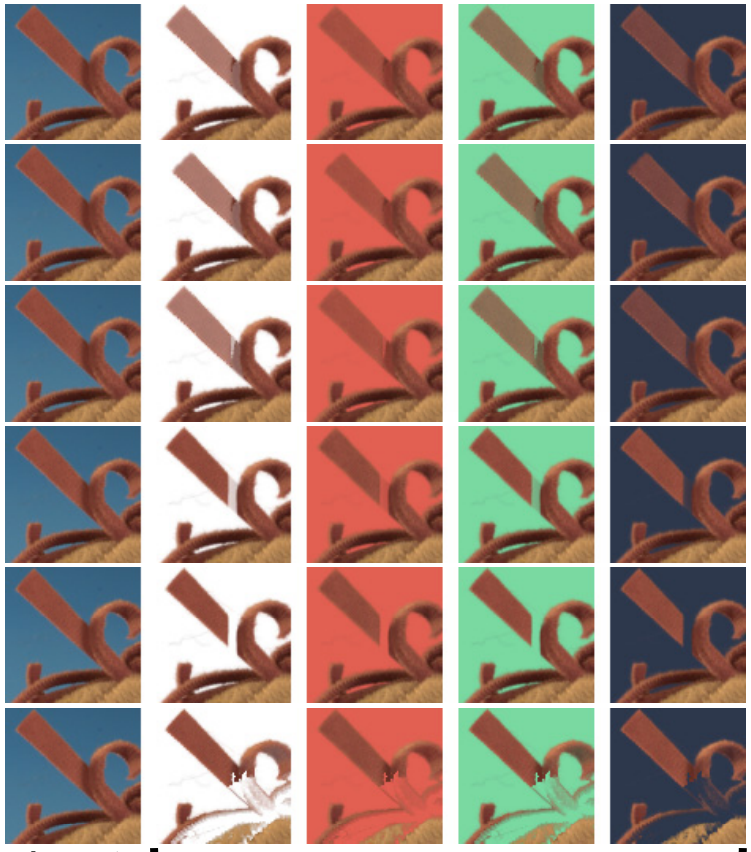
(a) GT alpha of Levin et al. '07 $\alpha \in [0.78,1]$

(b) GT alpha of our database $\alpha \in [0.78,1]$

This figure compares one example of the GT database of [Levin et al. '07] with one example of our database. Images (a) and (b) show the composite and part of the GT alpha for the two examples. Note that we only show alpha values between 0.78 and 1, which were then scaled to the full range of alpha values (therefore thin hairs with alpha below 0.78 are lost). In (a) a large number of opaque (true foreground) pixels are assigned to an alpha value much lower than 1, whereas (b) shows a much cleaner result.

User Study Results

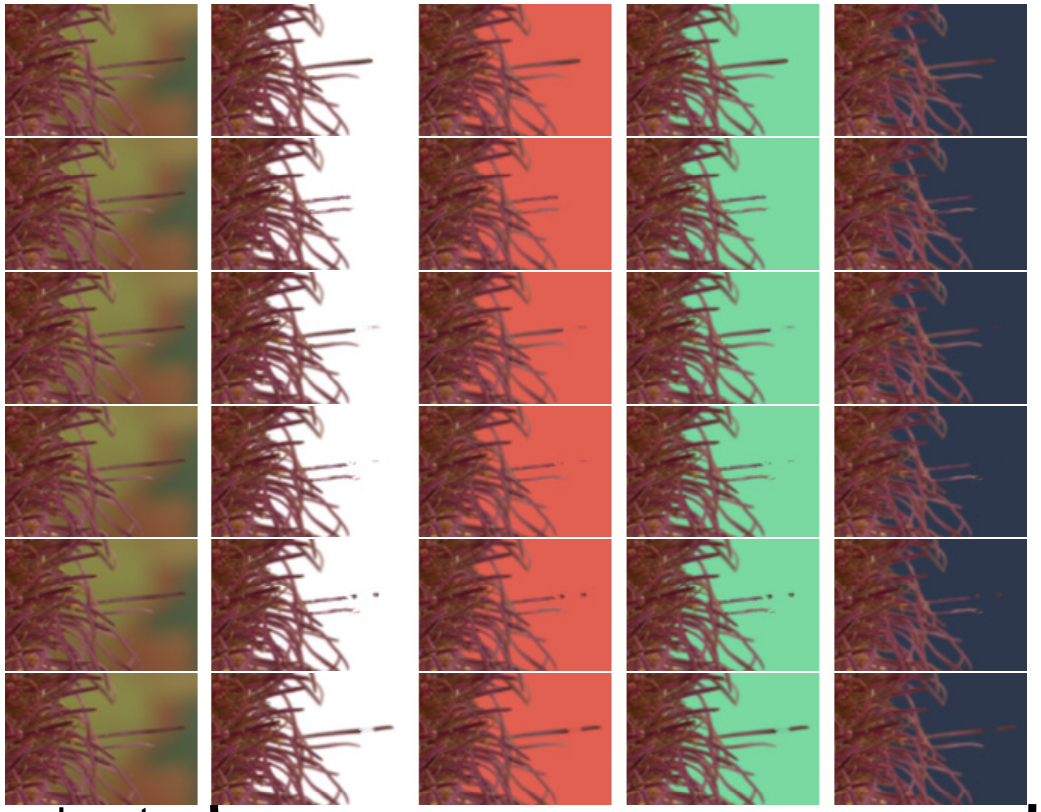
User Study Results – Connectivity (1)



Avg. User	Connectivity	MSE	SAD	Gradient
1	2 (8)	1 (0.7)	3 (232)	2 (40)
2	2 (8)	5 (1.9)	5 (312)	5 (82)
3	1 (4)	3 (0.8)	4 (243)	3 (42)
4	4 (2798)	2 (0.7)	1 (83)	1 (36)
5	5 (3827)	4 (1.3)	2 (111)	4 (76)
6	6 (18290)	6 (13.0)	6 (1211)	6 (203)

Left side: Study test case, showing compositions afflicted with connectivity artifacts. **Right side:** The corresponding rankings for the composites, derived from the average user and from four different error measures (i.e. Connectivity, MSE, SAD and Gradient). In brackets we show the corresponding absolute errors. We see that connectivity correlates quite well with the average user.

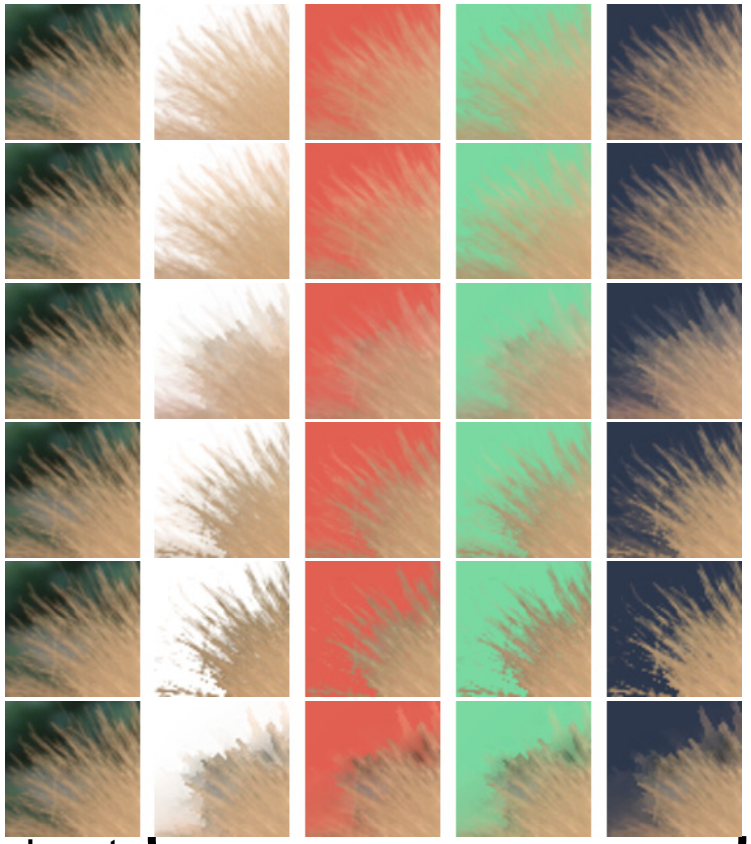
User Study Results – Connectivity (2)



Avg. User	Connectivity	MSE	SAD	Gradient
1	1 (6)	2 (0.19)	2 (53)	2 (15)
2	3 (51)	6 (0.57)	6 (128)	5 (57)
3	2 (16)	3 (0.37)	3 (82)	3 (37)
4	4 (68)	5 (0.55)	5 (122)	6 (58)
5	5 (173)	4 (0.45)	4 (112)	4 (49)
6	6 (182)	1 (0.11)	1 (22)	1 (13)

Left side: Study test case, showing compositions afflicted with connectivity artifacts. **Right side:** The corresponding rankings for the composites, derived from the average user and from four different error measures (i.e. Connectivity, MSE, SAD and Gradient). In brackets we show the corresponding absolute errors. We see that connectivity correlates quite well with the average user.

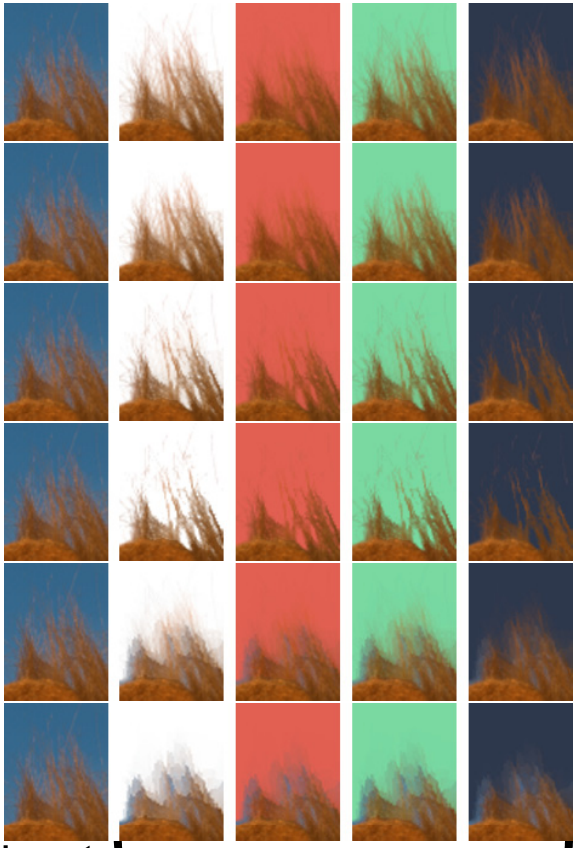
User Study Results - Gradient (1)



Avg. User	Gradient	MSE	SAD	Connectivity
1	2 (39)	3 (0.59)	3 (428)	2 (15)
2	1 (16)	1 (0.15)	1 (217)	1 (13)
3	3 (52)	2 (0.51)	2 (366)	2 (15)
4	4 (80)	4 (1.04)	4 (515)	5 (248)
5	5 (107)	6 (4.18)	6 (1030)	6 (506)
6	6 (118)	5 (2.05)	5 (730)	2 (15)

Left side: Study test case, showing compositions afflicted with gradient artifacts. **Right side:** The corresponding rankings for the composites, derived from the average user and from four different error measures (i.e. Gradient, MSE, SAD and Connectivity). In brackets we show the corresponding absolute errors. We see that our gradient measure correlates quite well with the average user.

User Study Results - Gradient (2)



Avg. User	Gradient	MSE	SAD	Connectivity
1	1 (2.1)	1 (0.17)	1 (276)	1 (0)
2	2 (2.8)	2 (0.41)	2 (469)	1 (0)
3	3 (29.4)	5 (2.07)	5 (909)	5 (23)
4	4 (44.3)	6 (3.70)	6 (1215)	6 (54)
5	5 (52.7)	3 (1.10)	3 (604)	1 (0)
6	6 (62.4)	4 (1.96)	4 (806)	1 (0)

Left side: Study test case, showing compositions afflicted with gradient artifacts. **Right side:** The corresponding rankings for the composites, derived from the average user and from four different error measures (i.e. Gradient, MSE, SAD and Connectivity). In brackets we show the corresponding absolute errors. We see that our gradient measure correlates quite well with the average user.

Evaluation Results

The following slides show examples from the evaluation. All results can be found at www.alphamatting.com

Performance on SAD/MSE (1)

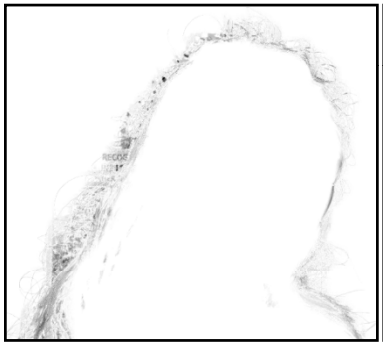
Input image

Closed-form

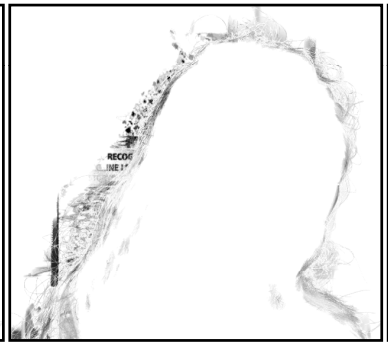
Robust matting

Random walk

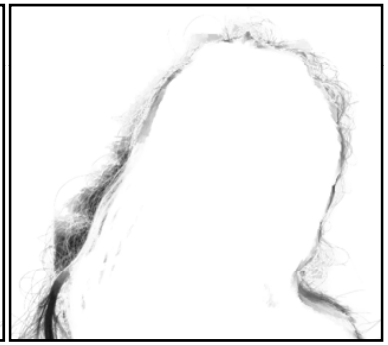
The input image shows highly textured background and color ambiguity.
Robust matting performs worse than Closed-form, due to large artifacts in the background, which are presumably due to erroneous color samples.
On the other hand we see that Robust matting helps to reveal more details in the hair (marked in red), which are oversmoothed by Closed-form and Random walk matting.



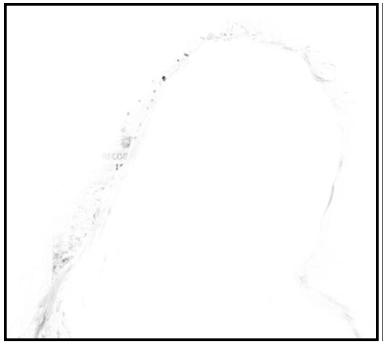
SAD; Rank: 1



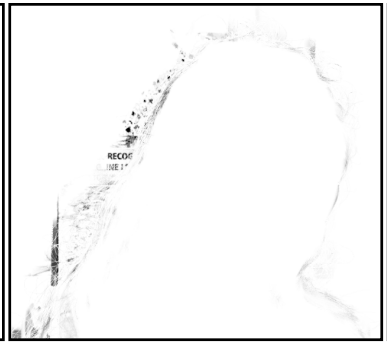
SAD; Rank: 2



SAD; Rank: 3



MSE; Rank: 1



MSE; Rank: 2

MSE; Rank: 3

Performance on SAD/MSE (2)

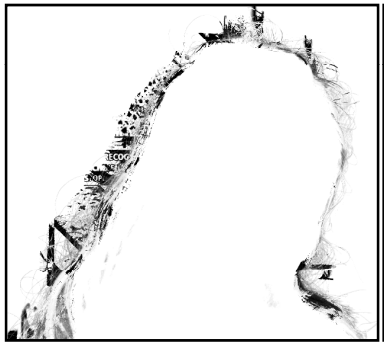
Input image

Easy matting

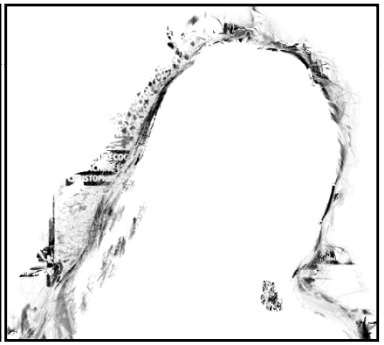
Bayesian matting

Poisson matting

The color model based approaches Easy matting and Bayesian matting rank worse than the pure propagation methods Closed-form matting and Random Walk matting. Similar to Robust Matting they show large erroneous regions due to the complex intensity variations in the image.
Poisson matting is the worst performer.

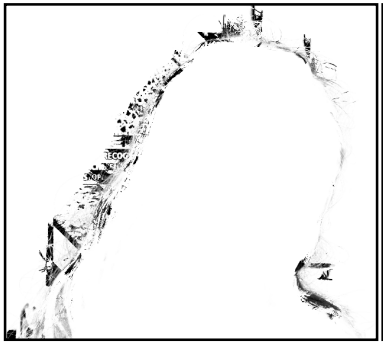


SAD; Rank: 4

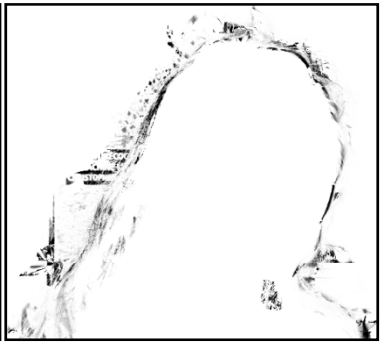


SAD; Rank: 5

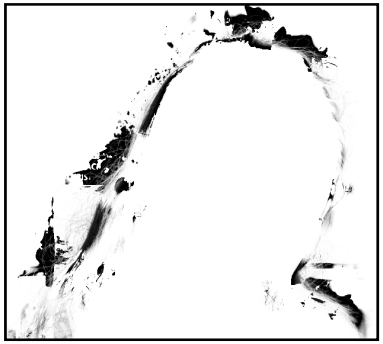
SAD; Rank: 6



MSE; Rank: 5



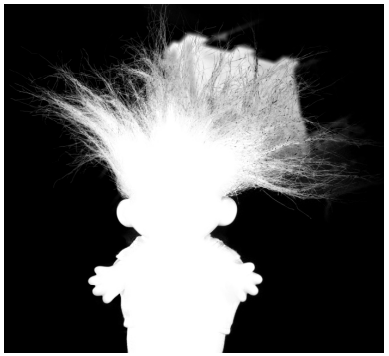
MSE; Rank: 4



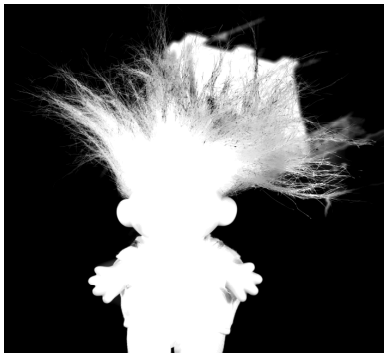
MSE; Rank: 6

Performance on Gradient error (1)

Input image



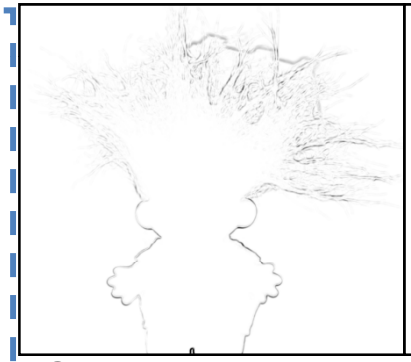
Closed-form



Robust matting

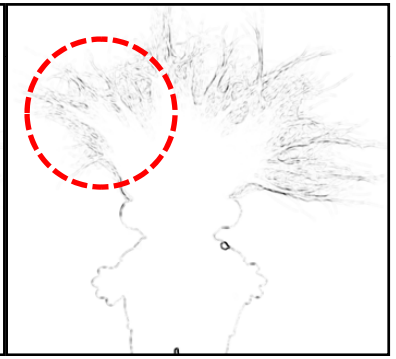
Random walk

The gradient measure emphasized artifacts of Random walk matting that are due to an oversmoothing in the hair region (marked in red) and widely ignores the large artifacts in the background (marked in green), generated by Robust matting and Closed-form matting. The background artifacts are heavily penalized by SAD/MSE and therefore Random walker matting performs worse under the Gradient measure, at least for this example.

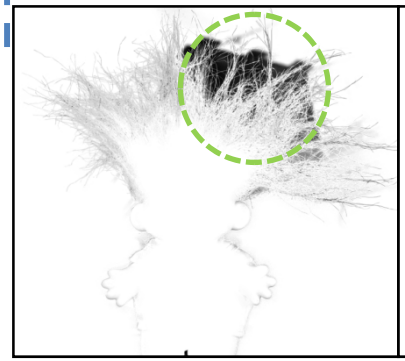


Gradient; Rank: 1

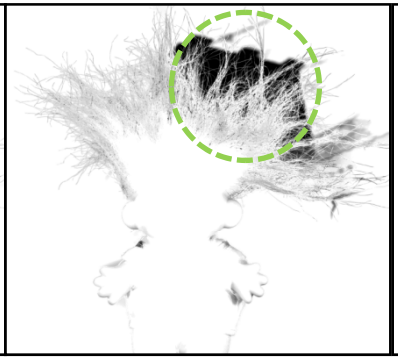
Gradient; Rank: 2



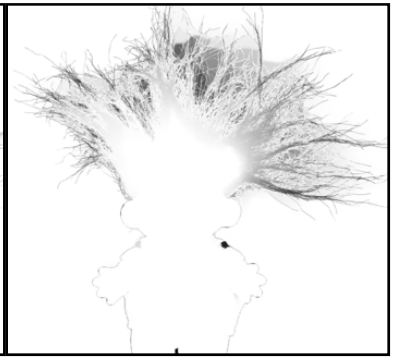
Gradient; Rank: 3



SAD; Rank: 2



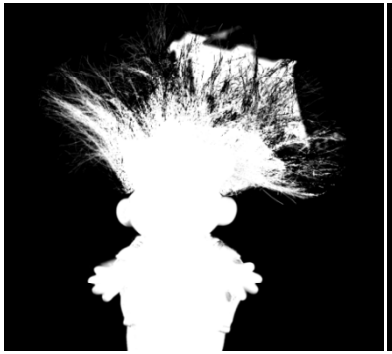
SAD; Rank: 3



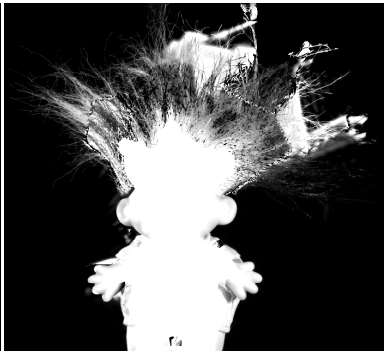
SAD; Rank: 1

Performance on Gradient error (2)

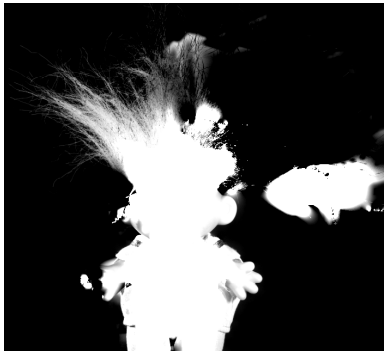
Input image



Easy Matting

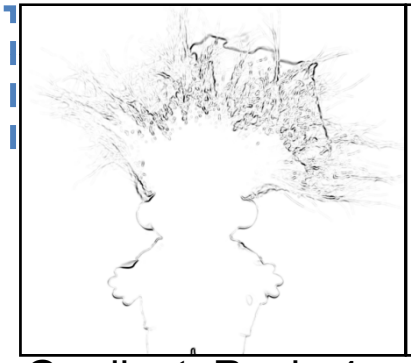


Bayesian matting

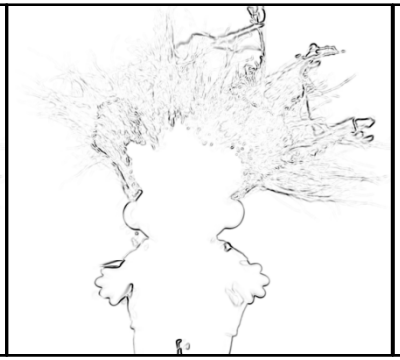


Poisson matting

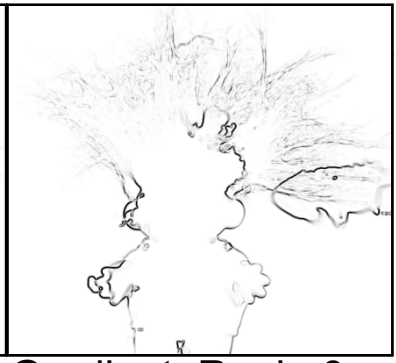
We see that Easy matting, Bayesian matting and Poisson matting perform similar under Gradient and SAD metric.



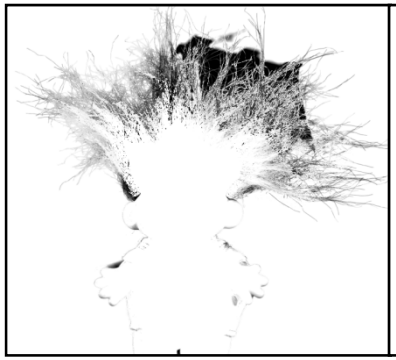
Gradient; Rank: 4



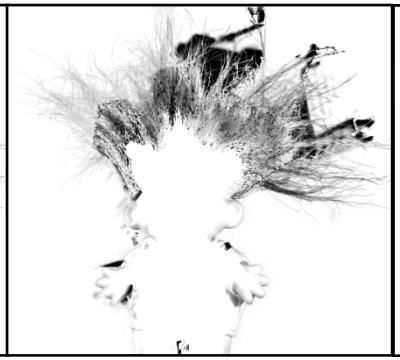
Gradient; Rank: 5



Gradient; Rank: 6



SAD; Rank: 4



SAD; Rank: 5

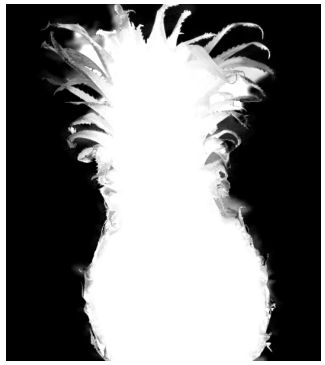
SAD; Rank: 6

Performance on Connectivity error (1)

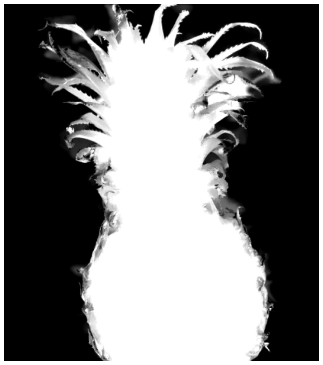
Input image

We see that the Random walk ranks first place, since its alpha matte is perfectly connected. However, under the SAD metric it only ranks 4th, which motivates algorithms that work well on all metrics.

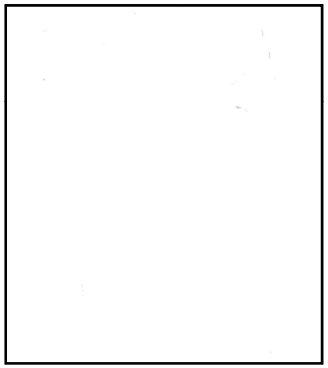
Random walk



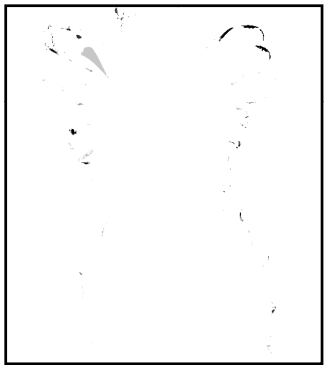
Closed-form



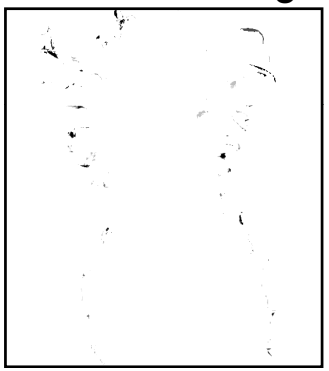
Robust matting



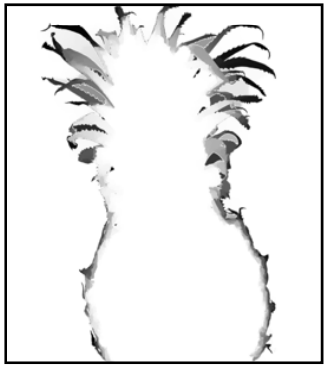
Connectivity; Rank: 1



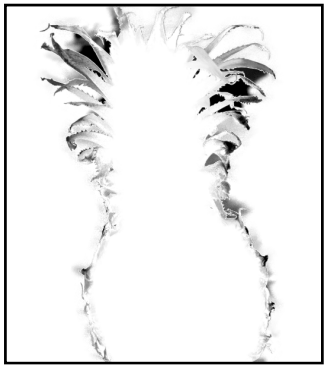
Connectivity; Rank: 2



Connectivity; Rank: 3



SAD; Rank: 4

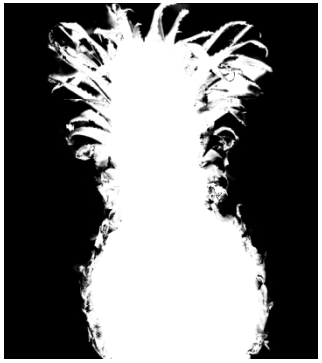


SAD; Rank: 2

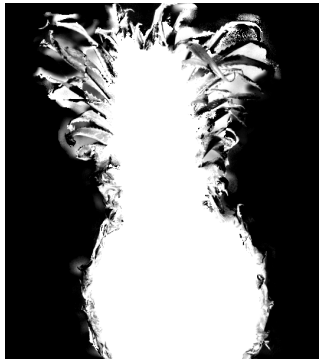
SAD; Rank: 1

Performance on Connectivity error (2)

Input image



Easy matting

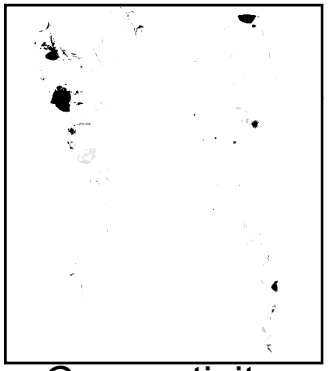


Bayesian matting

Poisson matting

Connectivity; Rank: 4

Connectivity; Rank: 5



Connectivity; Rank: 6

SAD; Rank: 3

SAD; Rank: 5

SAD; Rank: 6