Matting through Variational Inpainting

Kangyu Ni, Sheshadri Thiruvenkadam, and Tony Chan

Department of Mathematics, UCLA
{kni66,sheshad,chan}@math.ucla.edu

Abstract. While current *matting* algorithms work very well for some natural images, their performance is questionable in the presence of sharp discontinuities in the foreground and/or background regions. To resolve this problem, we propose to use variational PDE-based *inpainting* techniques to examine the matting problem, that are highly successful in inpainting geometric features into unknown regions.

Key words: matting, variational inpainting.

1 Introduction

Digital Image Composition is commonly used by the graphics community to create various scenarios for objects by extracting them from their original background and pasting them realistically to a new background. The above technique has been popularly used by the movie industry to transport images of actors captured in a controlled studio environment to new locations. A crucial step that precedes compositing images is to extract the object in question from its original background accurately (i.e. preserving the fractional nature of the object boundary). The above step is referred to as *Image Matting*, and additionally as in our work, if the background is unknown, it is referred to as *Natural Image Matting*.

Given an image $I : \Omega \rightarrow [0, 1]$ containing the object of interest, we wish to recover (α, F, B), using the commonly used *matting equation*:

$$I = \alpha F + (1 - \alpha)B.$$

Here F is the foreground intensity corresponding to the object, B is the background intensity, and α is the *soft*-segmentation (i.e. α-*matte*) of the object. Previous methods attempt to solve the above ill-posed problem by imposing statistical priors on (F, B) [4, 9], followed by regularity constraints [11] on α.

The problem is further simplified by a user specified *trimap* Fig.1, that partitions the image domain into three regions; definite foreground Ω_F, definite background Ω_B, and an unknown transition region D. Thus the problem of recovering (α, F, B) is restricted only to D. The above simplification approach means that the solution's quality *depends* on the initial-trimap accuracy. To generate an accurate trimap could be time-consuming. Recent works [5, 6, 12] attempt to resolve this problem by adding a prior region-growing step in the

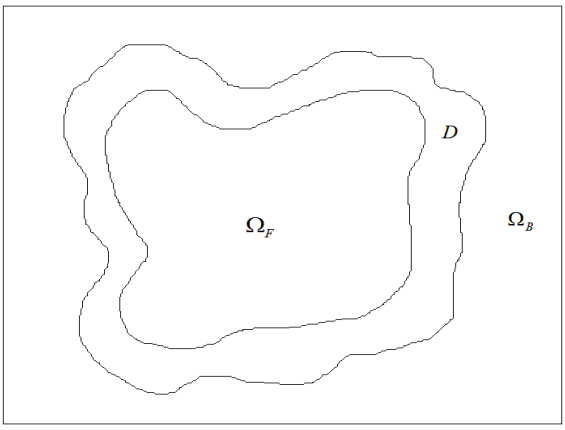

Fig. 1. Indications of subregions Ω_F, D, and Ω_B of the image domain Ω

(a) given image (b) desired result (c) by nearest values

Fig. 2. Example of interpolating the missing white region of an image by using the nearest known pixel values.

algorithm that computes a reasonable trimap from a user given rough guess. While current algorithms work very well for some natural images, their performance is questionable in the presence of sharp discontinues in the foreground and background regions. The reason is that these algorithms model the background and foreground intensities by statistical estimates of nearby pixel regions (usually the nearest neighbor). Thus they do not take into account the inherent geometry of the foreground and background regions. For example, in Fig.2, (a) is an image occluded by an object (white region). (b) shows the desired *inpainting* result. (c) is the result using nearest neighbor interpolation, variants of which are used in current matting algorithms such as Poisson matting [11]. We see from (c) that such methods are not able to achieve correct inpainting results. Hence, the extracted mattes may be erroneous even for images with simple geometric structures.

In response to the above problem, we focus on the class of non-texture images and propose to estimate the foreground and background by adapting existing non-texture variational inpainting methods, such as total variation (TV) inpainting [3] and Euler's elastica inpainting [1]. The former adapts the ROF image model [10] and is the first to take the Bayesian approach to inpainting problems. The latter improves TV inpainting and is a curvature-prior model. Both inpainting methods perform better in narrow inpainting regions than thick ones, see [2].

Typically, a matting algorithm consists of three iterative steps. We summarize our algorithm *vis-a-vis* these steps.

- **Trimap refinement:** A prior segmentation step is added to refine the approximate user defined trimap to conform to the actual α-transition region. We use an iterative scheme similar to Poisson matting, [11] to refine the trimap by thresholding the current iteration's α-estimate.
- **Extrapolating F and B:** The foreground and background intensities, F and B are extrapolated (i.e *inpainted*) into the transition region. We use PDE-based inpainting techniques (TV, elastica) that work well for interpolating geometric features.
- **Solving for α:** α is solved in the transition region using the matting equation subject to suitable priors. Similar to Poisson matting, we search for α in the Sobolev space $H^1(\Omega)$.

2 Extrapolating F and B

Here we describe the methods of extrapolating the background intensity $B : \Omega \to [0,1]$ and the foreground intensity $F : \Omega \to [0,1]$ through total variation (TV) and Euler's elastica inpainting. F and B are obtained by inpainting the data from the known regions, Ω_F and Ω_B into the unknown region D.

2.1 Total Variation Inpainting

TV inpainting is a PDE-based variational model and is adapted from the ROF denoising model [10]. It is based on the observation that edges play an important role in the geometry of an image. TV inpainting interpolates images across the missing regions, while preserving sharp edges. We propose to utilize this technique to extrapolate the background and foreground to examine the matting problem. The background is extrapolated the following energy minimization:

$$\min_{B \in BV(D \cup \Omega_B)} E_{tv}[B] = \int_{D \cup \Omega_B} |\nabla B|, \tag{1}$$

with constraint

$$B\,|_{\Omega_B} = I\,|_{\Omega_B}\ .$$

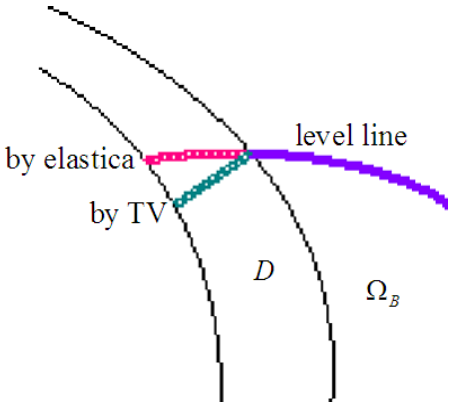

Fig. 3. Comparison of a level line to be completed in the unknown region by TV and elastica inpainting

Minimizing this energy functional is equivalent to connecting sharp edges according to the level sets in the known region. This can be seen by the coarea formula:

$$\int |\nabla B| dx = \int_0^1 \int_{\Gamma_\lambda} ds d\lambda \, ,$$

where $\Gamma_\lambda = \{x : B(x) = \lambda\}$ is the level set and ds is the arc length of the level sets.
The gradient descent of the Euler-Lagrange equation (1) is

$$\frac{\partial B}{\partial t} = 1_D \nabla \cdot (\frac{\nabla B}{|\nabla B|}), \tag{2}$$

with condition

$$B \mid_{\Omega_B} = I \mid_{\Omega_B} .$$

The boundary condition along the boundary between D and Ω_F is

$$\frac{\partial B}{\partial \overrightarrow{\nu}} = 0.$$

The formulation for estimating the foreground is similar.

2.2 Euler's Elastica Inpainting

TV inpainting does not always connect edges correctly. For example, when the inpainting region is too large, the level lines may form corners at the boundary of the inpainting region, in order to achieve the smallest distance connection.

This usually does not agree with visual perception. The geometry of inpainting regions is crucial for TV inpainting.

Euler's elastica inpainting improves TV inpainting by additionally penalizing the curvature. As a result, the level lines are extended properly into the inpainting region. Figure 3 is an example of completing a level line inside the unknown region according to TV and elastica inpainting, respectively. The estimated background through Euler's elastica inpainting is obtained by the following:

$$\min_{B\in BV(D\cup\Omega_B)} E_{elas}[B] = \int_{D\cup\Omega_B} (a+b\kappa^2)|\nabla B|dx, \tag{3}$$

with condition

$$B\,|_{\Omega_B} = I\,|_{\Omega_B}\ .$$

In the functional, a and b are positive constants and $\kappa = \nabla\cdot(\frac{\nabla B}{|\nabla B|})$ is the curvature of u. The weak absolute curvature is defined for BV functions in [1]. Minimizing this energy functional is equivalent to connecting sharp edges according to the curvature of the level sets in the known region. This again can be seen by the coarea formula:

$$\int (a+b\kappa^2)|\nabla B|dx = \int_0^1 \int_{\Gamma_\lambda} (a+b\kappa^2)dsd\lambda\ .$$

The gradient descent of the Euler-Lagrange equation of (3) is

$$\frac{\partial B}{\partial t} = 1_D \nabla\cdot\overrightarrow{V} \tag{4}$$

with

$$B\,|_{\Omega_B} = I\,|_{\Omega_B},$$

where

$$\overrightarrow{V} = \phi(k)\overrightarrow{n} - \frac{\overrightarrow{t}}{|\nabla B|}\frac{\partial\phi'(k)|\nabla B|}{\partial\overrightarrow{t}} \tag{5}$$

$$\text{and } \overrightarrow{n} = \frac{\nabla B}{|\nabla B|}\ ,\ \overrightarrow{t} = \overrightarrow{n}^{\perp}\ \ ,\ \ \frac{\partial}{\partial\overrightarrow{t}} = \overrightarrow{t}\cdot\nabla.$$

The boundary conditions along the boundary between D and Ω_F is

$$\frac{\partial B}{\partial\overrightarrow{\nu}} = 0 \text{ and } \frac{\partial\phi'(\kappa)|\nabla B|}{\partial\overrightarrow{\nu}} = 0. \tag{6}$$

3 Matte Extraction and Trimap Refinement

Given F and B, the matte α is estimated by minimizing the following energy:

$$\min_{\alpha}\int_D (\alpha F + (1-\alpha)B - I)^2 + \frac{\lambda}{2}|\nabla\alpha|^2 dx\ , \tag{7}$$

with constraints

$$\alpha\,|_{\Omega_F} = 1 \text{ and } \alpha\,|_{\Omega_B} = 0.$$

The first term in the energy seeks an α that adheres to the matting equation. The second term enforces the smoothness of α, according to parameter λ. The constraints enforce α to equal 1 in the definite foreground region and equal 0 in the definite background region. The minimized α is then thresholded to update the trimap for the subsequent iteration.

4 Numerical Method

In this section, we describe the numerical method of our iterative scheme, $n = 1, 2, \ldots$.

1. **Trimap refinement:**

$$\Omega_F^{(n)} = \{x \in \Omega : \alpha^{(n-1)}(x) > 0.97\}$$

$$\Omega_B^{(n)} = \{x \in \Omega : \alpha^{(n-1)}(x) < 0.03\}$$

2. **Extrapolating F and B:**

$$B^{(n)} = \arg\min_B E[B], \text{ with } B|_{\Omega_B^{(n)}} = I|_{\Omega_B^{(n)}}$$

$$F^{(n)} = \arg\min_F E[F], \text{ with } F|_{\Omega_F^{(n)}} = I|_{\Omega_F^{(n)}}$$

3. **Solving for α:**

$$\alpha^{(n)} = \arg\min_\alpha \int (\alpha F^{(n)} + (1-\alpha)B^{(n)} - I)^2 + \frac{\lambda}{2}|\nabla\alpha|^2,$$

with constraint $\alpha|_{\Omega_F^{(n)}} = 1$ and $\alpha|_{\Omega_B^{(n)}} = 0$

$E[\cdot]$ can be either $E_{tv}[\cdot]$ or $E_{elas}[\cdot]$. Initially, $\alpha^{(0)}$ is given by the user. Repeat step $1-3$ until α converges, i.e. $d(\alpha^{(n)}, \alpha^{(n+1)}) < \epsilon$, where d for example is the l_1-norm and ϵ is small.

In the second step, TV inpainting is implemented by lagged diffusivity and fixed point iteration described in [3] for (2). The numerical scheme is

$$B_{i,j}^{n+1} = \left(\frac{B_{i+1,j}^n}{|B_{i+\frac{1}{2},j}^n|} + \frac{B_{i,j+1}^n}{|B_{i,j+\frac{1}{2}}^n|} + \frac{B_{i-1,j}^n}{|B_{i-\frac{1}{2},j}^n|} + \frac{B_{i,j-1}^n}{|B_{i,j-\frac{1}{2}}^n|}\right)/A,$$

where

$$A = \frac{1}{|B_{i+\frac{1}{2},j}^n|} + \frac{1}{|B_{i,j+\frac{1}{2}}^n|} + \frac{1}{|B_{i-\frac{1}{2},j}^n|} + \frac{1}{|B_{i,j-\frac{1}{2}}^n|}.$$

For elastica inpainting, to avoid modeling boundary conditions (6), we extend the inpainting region of B from D to $D \cup \Omega_F$ and extend the inpainting region

of F from D to $D \cup \Omega_B$. The inpainting result of B (resp. F) on Ω_F (resp. Ω_B) is less important since it is not utilized in optimizing α. We use the numerical scheme described in [1], which is an explicit scheme for

$$\frac{\partial B}{\partial t} = 1_{D \cup \Omega_F} |\nabla B| \nabla \cdot \overrightarrow{V}. \tag{8}$$

As suggested in [8], the factor $|\nabla B|$ accelerates the original time marching equation (4) and is discretized by central differencing. The term $\nabla \cdot \boldsymbol{V}$ is discretized by half-point central differencing,

$$\nabla \cdot \boldsymbol{V}_{i,j} = (V^1_{i+\frac{1}{2},j} - V^1_{i-\frac{1}{2},j}) + (V^2_{i,j+\frac{1}{2}} - V^2_{i,j-\frac{1}{2}}),$$

where $(V^1, V^2) = \boldsymbol{V}$ in (5).

In (5), the discretization of D_x and D_y at the x-half-point are

$$D_x B_{i+\frac{1}{2},j} = \frac{1}{2}(B_{i+1,j} - B_{i,j}) \text{ and}$$

$$D_y B_{i+\frac{1}{2},j} = \text{minmod}(\frac{1}{2}(B_{i+1,j+1} - B_{i+1,j-1}), \frac{1}{2}(B_{i,j+1} - B_{i,j-1})),$$

where

$$\text{minmod}(a, b) = \frac{\text{sgn}(a) + \text{sgn}(b)}{2} \min(|a|, |b|).$$

The discretizations at other half-points are similar. The reader may consult the details in [1].

The solution of α matte is obtained by the steepest descent of Euler-Lagrange equation for (7)

$$\alpha^{n+1} = \alpha^n - \delta t([\alpha^n F + (1 - \alpha^n)B - I](F - B) - \lambda \triangle \alpha^n),$$

where

$$\triangle \alpha = -4\alpha_{i,j} + \alpha_{i+1,j} + \alpha_{i-1,j} + \alpha_{i,j+1} + \alpha_{i,j-1}$$

is the usual five point discrete Laplacian.

5 Experimental Results

In this section, we show experimental results on both synthetic and real images, and compare our algorithm with a method that uses nearest neighbor inpainting. The experiments indicate that our method outperforms such nearest neighbor based matting methods, especially for images with sharp edges. Specifically, the *comparison method* we use here is similar to Poisson matting; models α in $H^1(\Omega)$ and iteratively refines the trimap region, and uses nearest neighbor interpolation to fill in F and B values into the unknown region.

The first example is shown in Fig.4. The foreground of the given image (a) is a constant and the background (d) is a bar. The given trimap (b) indicates the definite foreground, background, and unknown regions by $\alpha = 1$, $\alpha = 0$, and $\alpha = 0.5$, respectively. The first column are the ground truth matte (c), the extracted mattes by elastica inpainting (e), TV inpainting (g), and the comparison method (i). The respective mattes are (d), (f), (h), and (j). The background (f) obtained by the elastica inpainting is very close to the ground truth background (d). As a result, the corresponding matte (e) is also very close to the ground truth matte (c). The performance of the TV inpainting strongly depends on the geometry of the inpainting region and the image data. As expected, the TV inpainting is able to recover the underlying geometry at the region with smaller width and fails at the larger scale region. The corresponding matte (g) also reflects the accuracy of the extrapolated background (h). The comparison method is not able to recover the underlying geometry (j) and the extracted matte (i) is erroneous.

The second example is shown in Fig.5. The given image (a) is a starfish, as the foreground, and the background has stripes occluded by the starfish. The second row shows the elastica inpainting method (c) is able to extract the α matte accurately while the result of the comparison method (d) is erroneous near sharp gradients of the background. The third row shows the resulting foreground by elastica inpainting (e) and nearest neighbor (f). The unknown region is indicated in between the red curves. The last row shows the resulting background by elastica inpainting (g) and nearest neighbor (h). Elastica inpainting outperforms nearest neighbor thus our method is more efficient in extracting the α matte.

The third example is shown in Fig.6. The object of the given image (a) is has more interesting boundaries and the background has a structure occluded by the foreground. The second and third rows show the extracted mattes and extracted foreground by the elastica inpainting and nearest neighbor, respectively. Observe that near the boundary of the bear and bar, the comparison method (f) is erroneous. This can be seen clearly in the fourth row, which shows the selected local regions of elastica inpainting (g), the comparison method (h), and the original image (i). The last two rows show the extrapolated foreground and background, respectively. The unknown region is indicated in between the red curves.

The above results demonstrate that when the background has geometric structures occluded by the foreground, curvature-prior models, such as elastica inpainting, are able to recover the background correctly. Consequently, the extracted matte is more accurate. From our experiment, the extracted mattes from the iterative scheme do not change significantly after about 3 iterations. The numerical error converges in about 30 iterations.

6 Conclusion and Future Work

In this paper, we propose to employ PDE-based variational inpainting methods to examine the matting problem. Our results show that elastica inpainting is effective for non-texture images. As future work we would like to improve

our matting algorithm in two aspects. The first is to speed up the algorithm. In particular, the elastica inpainting in the second step involves solving a stiff fourth-order PDE, thus the process is computationally expensive. Recently, [7] proposed a method that approximates and changes the elastica inpainting solution to a combinatorial optimization problem. By their experiment, this method is two orders of magnitude faster than the stiff PDE solver. The second is to add a region-growing step to generate a reasonable trimap from the user's rough guess.

7 Acknowledgements

This research is supported by ONR grant N00014-06-1-0345 and NSF grant DMS-0610079.

References

1. T. Chan, S. H. Kang and J. Shen, Euler's Elastica and Curvature Based Inpaintings, *SIAM J. Appl. Math.*, 63:564-594, 2002.
2. T. Chan and S. H. Kang, An Error Analysis on Image Inpainting Problems, *J. Math. Imag. Vision*, to appear.
3. T. Chan and J. Shen, Mathematical Models for Local Nontexture Inpaintings, *SIAM J. Visual Comm. Image Rep.*, 12:436-449, 2001
4. Y. Y. Chuang, B. Curless, D. H. Salesin and R, Szeliski, A Bayesian Approach to Digital Matting, *Proceeding of CVPR 2001*, Vol.II, 264-271.
5. Y. Guan, W. Chen, X. Liang, Z. Ding, and Q. Peng, Easy Matting-A Stroke Based Approach for Continuous Image Matting, *Eurographics 2006*, Vol 25(2006), Number 3.
6. O. Juan and R. Keriven, Trimap Segmentation for Fast and User-Friendly Alpha Matting, *VLSM 2005*, pp.186-197.
7. K. Ni, D. Roble and T. Chan, A Texture Synthesis Approach to Elastica Inpainting, ACM SIGGRAPH 2007 sketch program.
8. A. Marquina and S. Osher, *Lecture Notes in Computer Science*, volume 1682m chapter "A new time dependent model based on level set motion for nonlinear deblurring and noise removal", pp. 429-434, 1999.
9. M. A. Ruzon and C. Tomasi, Alpha Estimation in Natural Images, *Proceeding of CVPR 2000*, 18-25.
10. L. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Phys. D*, 60:259-268, 1992.
11. J. Sun, H. Jia, C. Tang and H. Shum, Poisson Matting, Proc. of ACM SIGGRAPH, pp.315-321, 2004.
12. J. Wang, M. F. Cohen, An Iterative Optimization Approach for unified image segmentaion and matting, In Proceedings of International Conference on Computer Vision (2005), pp.936-943.

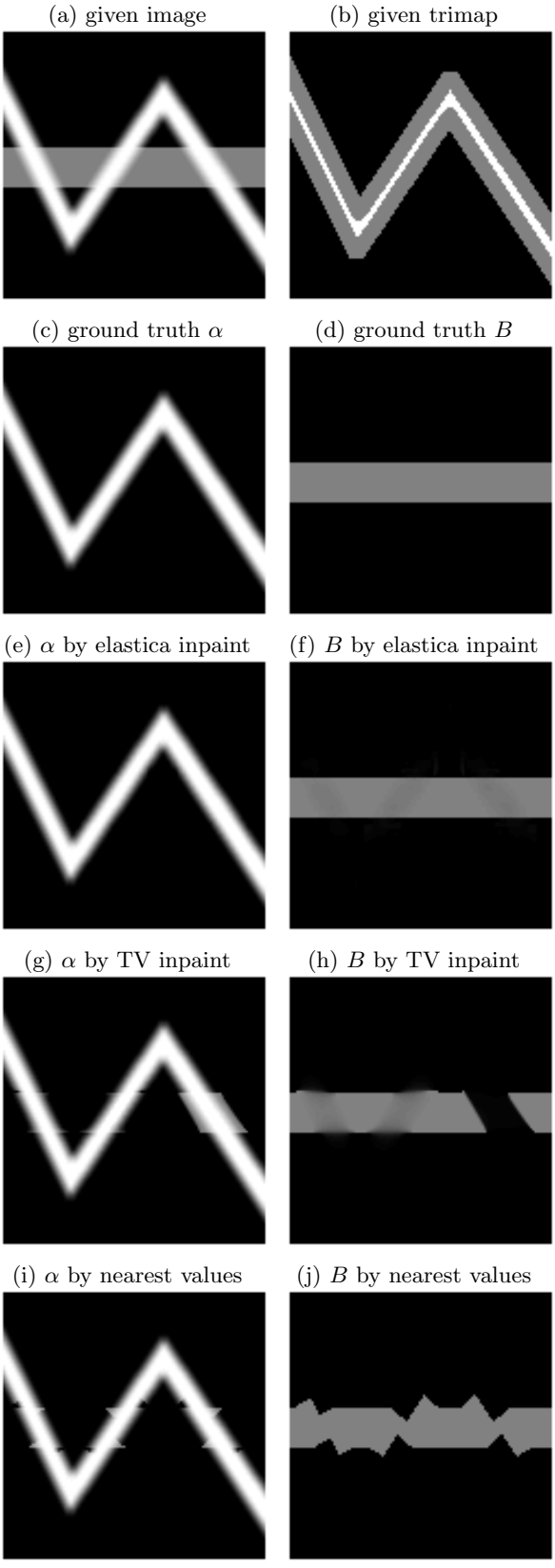

Fig. 4. Comparison of the Euler's elastica inpainting, TV inpainting, and nearest values for extrapolating the background and the corresponding extracted mattes.

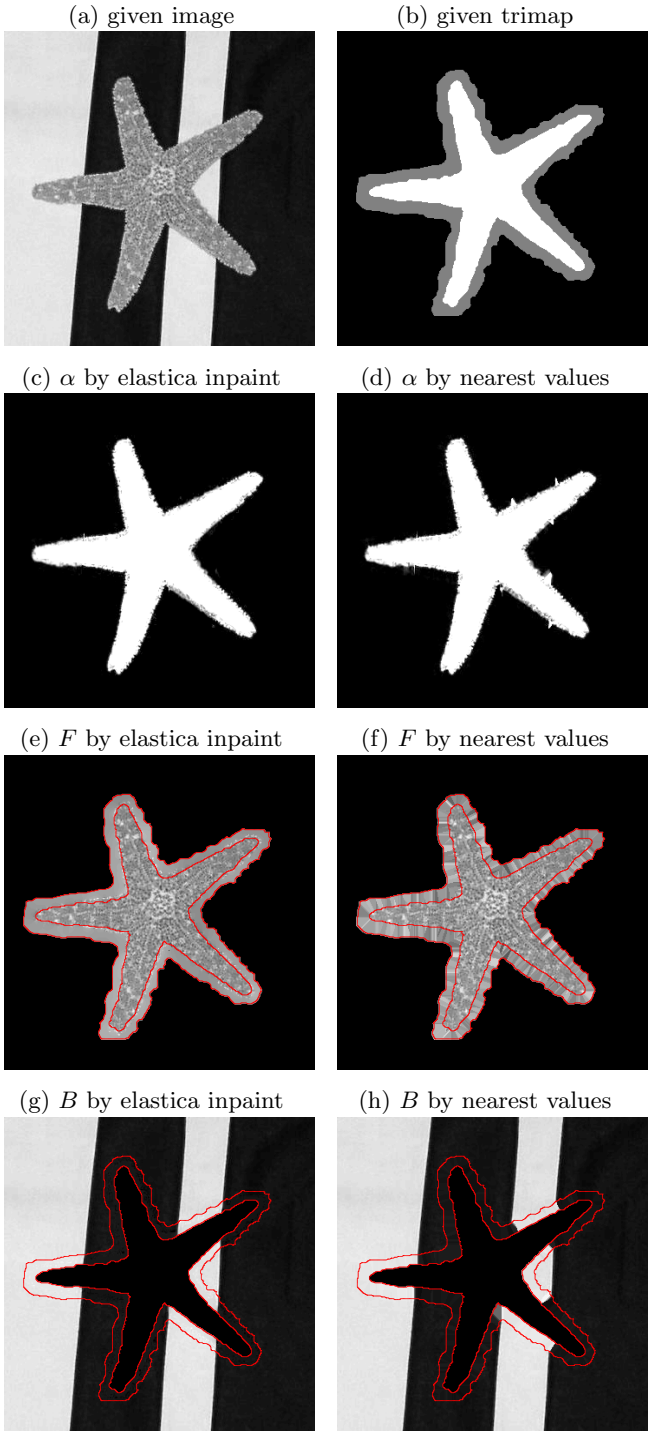

Fig. 5. The proposed method by utilizing the Euler's elastica inpainting produces a satisfying matte of the starfish.

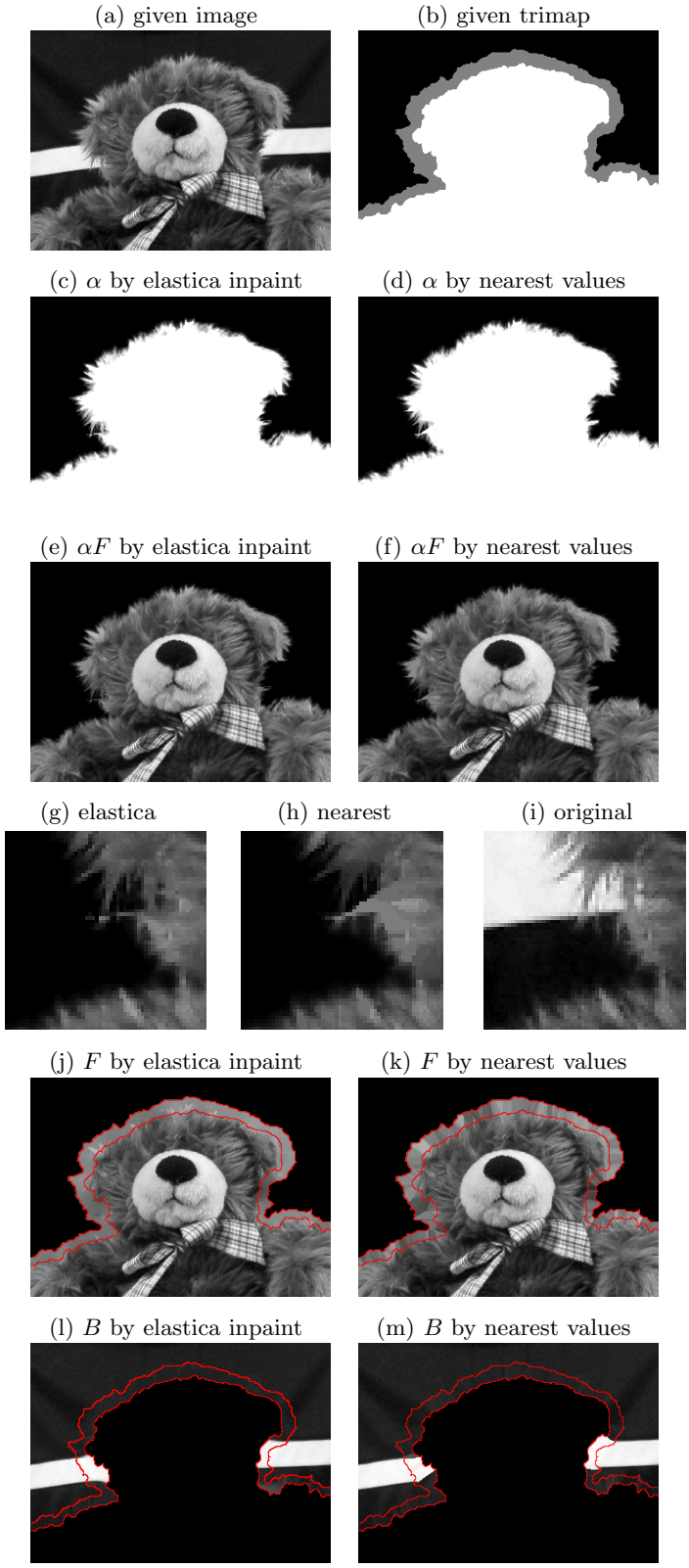

Fig. 6. The proposed method by utilizing the Euler's elastica inpainting produces a satisfying matte of the bear.