

VARIATIONAL APPROACH TO NATURAL IMAGE MATTING*

SHESHADRI. R. THIRUVENKADAM[†], TONY. F. CHAN[‡], AND BYUNG-WOO HONG[§]

Abstract. A variational PDE based model is presented for natural image matting. Given an input image with a trimap guess, we search for BV solutions for the α - *matte*, and the foreground/background intensities. Additionally, regularity conditions that comes naturally from the image formation model are imposed for the α - *matte*. Our model has two advantages. Firstly, the geometry based TV regularization is better (for non texture images with sharp edges) at inpainting F and B from known regions than nearest neighbour based approaches (used by many matting algorithms). Secondly, the additional conditions on α gives our model robustness with respect to the initial trimap guess (another issue in current algorithms). We demonstrate our approach by experiments on real and synthetic data.

1. Introduction. *Digital Image Composition* is commonly used by the graphics community to create various scenario for objects by extracting them from their original background scene and pasting them realistically to new background scenes. The above technique has been popularly used by the movie industry to transport images of actors captured in a controlled studio environment to novel locations [5]. A crucial step that precedes compositing images is extracting the object in question from its original background scene realistically (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, referred to as *Natural Image Matting*.

Given an image $I : \Omega \rightarrow \mathbb{R}$ containing the object of interest, natural image matting aims to recover (α, F, B) , where F and B are the foreground and background intensities, and α is the *soft*-segmentation (i.e. α -*matte*) of the object that reflects the proportion of foreground and background [13]. The image I is related to (α, F, B) through the commonly used matting equation $I = \alpha F + (1 - \alpha)B$. Recovery of (α, F, B) from such an I is an inherently ill-posed problem that requires further constraints. A variety of matting algorithms have been introduced using different number of images and different *a priori* assumptions on F, B and α .

One of the initial approaches for image matting is blue screen matting where F and α are to be estimated while B is a known constant [16]. The known constant background condition simplifies the problem, yet is still insufficient to fully constrain the problem [12]. Further such an assumption on the background is generally not applicable to natural images, and hence requires a controlled environment or imaging scenario. Recently, a number of natural image matting algorithms have been proposed to estimate the alpha matte from a single image based on a manually partitioned *trimap* that consists of definitely foreground F , definitely background B , and unknown regions where statistics of intensity distributions of the known regions are propagated into the unknown regions [2, 15, 8, 4, 17, 1]. In Knockout [2], F and B are assumed to be smooth and the estimated α is obtained as a weighted average of F and B . Instead of simple weighted sum as the estimation of a local intensity distribution, a mixture of un-oriented Gaussians has been used in [15] and principal component analysis (PCA) has been used in [8]. In [4], Chuang et al. proposed Bayesian matting algorithm where a mixture of oriented Gaussians is employed to estimate the local

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

[†]Department of Mathematics, University of California, Los Angeles, sheshad@math.ucla.edu.

[‡]Professor of Mathematics, University of California Los Angeles, and Assistant Director, Directorate for Mathematics & Physical Sciences, The National Science Foundation, chan@math.ucla.edu.

[§]Computer Science Department, University of California, Los Angeles, hong@cs.ucla.edu.

distribution and F , B and α are estimated by a maximum a posterior (MAP) in a Bayesian framework. The Poisson matting algorithm was proposed in [17], where the alpha matte is obtained from its gradient field by solving a Poisson equation using the boundary information provided by a trimap. These methods generally produce good results, however the major drawback is their considerable dependency on the initial condition that is given by a user specified trimap. For a good matte, the unknown region in the trimap has to be as narrow as possible. However, it is intractable to provide optimal and consistent trimap manually.

This naturally leads to introduce the combination of image segmentation and matting as in [3, 11, 14, 10] where a hard binary segmentation based on graph-cut framework is initially performed followed by estimating a fractional α matte from the segmentation results. Another approach is based on a few user-specified strokes rather than trimap using Belief Propagation techniques [19] and Markov Random Field (MRF) [6] for modeling unknown regions. These methods generally involve computationally expensive iterations in the optimization of non-linear systems that might converge to local minima. There are also several works that incorporate temporal information or optical flow using multiple images of the object of interest in order to constrain the matting equation in [20, 21, 7, 18].

Even though one can see excellent matte results in recent works, we believe that the root ill-posed nature of the matting problem is still not completely tackled, which primarily contributes to stability issues such as the above trimap dependency problem. To counter the above mentioned issues, the main focus of our paper is to *properly* regularize the recovery of α , F , and B . We make two simplifying assumptions which allows us to choose appropriate regularizers for α , F , and B . Firstly the images that we deal with are non-texture, thus we model F and B as functions in BV space. Geometry based (TV) regularization has been used before for matting [9], and does not give artifacts in the matte, especially for images with sharp edges. Secondly we assume that the foreground object is in focus, and we regularize α in a manner consistent with such an image formation model. The above choice of α -regularization is primarily responsible for the stability of computed solutions w.r.t the initial trimap guess and is demonstrated within our experiments. The usual way of regularizing just the gradient of α (used by many matting algorithms) does not guarantee stability of the computed solution, and can result in a noisy matte even for simple images with sharp edges.

Briefly, in this work, we first consider the image formation model for an object in focus. This is easily shown to give the commonly used matting equation. The matte α is seen to be the fractional area of the object region within a square shaped region of fixed size (a camera parameter). This form of α results in some necessary conditions which an α solution has to satisfy. We formulate a variational energy that uses a search space $BV(\Omega)$ for α , F , and B , additionally constraining α to satisfy the above conditions.

2. Image Formation Model. Suppose that O is the object of interest assumed to be *in focus* in a 3D scene. Let $I : \Omega \rightarrow \mathbb{R}$ represent the image of the above scene. Since the object O is in focus, we can approximate the projection of O by a binary function u defined on Ω . Then the intensity recorded at each point x after considering blurring effects due to pixel averaging and camera optics is given by:

$$I(x) = \frac{1}{4\epsilon^2} \int_{S_x} (u(t)f(t) + (1 - u(t))b(t)) \eta(x - t) dt. \quad (2.1)$$

η is the blurring function with compact support within S_0 . Here, S_x denotes a square shaped region centered at x and width 2ϵ , $\epsilon > 0$. f and b are the observed intensities of the object and the background scene. The above equation can further be simplified into the commonly used *matting equation*

$$I = \alpha F + (1 - \alpha)B \quad (2.2)$$

Here $\alpha(x)$ is the fractional area of the object-region within S_x ,

$$\alpha(x) = \frac{1}{4\epsilon^2} \int_{S_x} u \, dt,$$

Henceforth, we will refer to such functions as *transition functions*. Also,

$$F(x) = \frac{\int_{S_x} u f \eta \, dt}{\int_{S_x} u \, dt}, \text{ and } B(x) = \frac{\int_{S_x} (1 - u) b \eta \, dt}{\int_{S_x} 1 - u \, dt}$$

whenever well-defined, are the average intensities of the object and the background recorded within S_x .

Thus given an image I as described above with additive gaussian noise, we wish to recover α , F and B . Once α , F and B are solved for, we can combine α and F with a novel background $\hat{B} = \hat{b} * \eta$ to form the composite image,

$$\begin{aligned} \hat{I} &= \alpha F + (1 - \alpha)\hat{B} \\ &\approx \frac{1}{4\epsilon^2} \int_{S_x} (u(t)f(t) + (1 - u(t))\hat{b}(t)) \eta(x - t) \, dt. \end{aligned}$$

The above composition is *realistic* in the sense that it is close to the image of the object O in the novel background, assuming similar lighting conditions.

We use a variational framework to formulate the above inverse problem along with appropriate regularization terms for the unknowns. It is to be noted that α within the matting equation (2.2) takes the special form of a transition function, i.e. $\alpha(x) = \frac{1}{4\epsilon^2} \int_{S_x} u \, dt$. In contrast to current matting models that regularize only $\nabla\alpha$, we specifically search for α -solutions that are transition functions. We have found that such a restrictive search space is essential for the stability of computed solutions. Since α is related to some object region u , one possibility is to directly search for a solution u , that gave rise to I in (2.2). But this is computationally challenging, since one has to resolve u to within subpixel accuracy for natural images. Also, it seems to be unnecessary for the problem, since all that is needed to composite the foreground object to a different background scene is the matte, α . As a result, we will search for an α -solution in a suitable space, additionally constrained by some necessary conditions which a transition function has to satisfy.

2.1. Necessary Conditions for a α -solution. Here we look at some of the properties that an α estimate has to necessarily satisfy. We further assume that the object region defined by u has finite perimeter, thus $u : \Omega \rightarrow \{0, 1\}$ is in the space $BV(\Omega)$. Firstly we note that $\alpha = \frac{1}{4\epsilon^2} \int_S u \, dt$ is in the Sobolev space $H^1(\Omega)$.

To motivate the discussion, we will first consider the 1D case, i.e $\Omega \subset \mathbb{R}$ and $\alpha(x) = \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} u(y)dy$. We see that $\alpha'(x) = \frac{1}{2\epsilon}[u(x+\epsilon) - u(x-\epsilon)]$. Thus

$$\alpha' \in BV(\Omega) \text{ and } \alpha' \in \{0, \pm \frac{1}{2\epsilon}\}. \quad (2.3)$$

Hence for the 1D matting problem, we search for α in $H^1(\Omega)$, with additional constraints on α' given above.

For the 2D case, we have $\alpha(x, y) = \frac{1}{4\epsilon^2} \int_{x-\epsilon}^{x+\epsilon} \int_{y-\epsilon}^{y+\epsilon} u(s, t) dsdt$. From the form of α , it is straightforward to show that there exists a band (*transition region D*) of width 2ϵ , where

$$\begin{aligned} 0 < \alpha(x) < 1, \quad x \in D, \text{ and} \\ \alpha(x) \in \{0, 1\}, \quad x \in D^c \end{aligned} \quad (2.4)$$

Differentiating α gives $\alpha_x(x, y) = \frac{1}{4\epsilon^2} \int_{y-\epsilon}^{y+\epsilon} u(x+\epsilon, t) - u(x-\epsilon, t) dt$, and $\alpha_y(x, y) = \frac{1}{4\epsilon^2} \int_{x-\epsilon}^{x+\epsilon} u(s, y+\epsilon) - u(s, y-\epsilon) ds$. A necessary bound

$$|\nabla\alpha| \leq \frac{\sqrt{2}}{2\epsilon} \quad (2.5)$$

follows from the above equations. We further notice from $\alpha_{xy}(x, y) = \frac{1}{4\epsilon^2}[u(x+\epsilon, y+\epsilon) - u(x-\epsilon, y+\epsilon) - u(x+\epsilon, y-\epsilon) + u(x-\epsilon, y-\epsilon)]$ that

$$\begin{aligned} \alpha_{xy} \in BV(\Omega) \text{ and specifically} \\ \alpha_{xy} \in \{0, \pm \frac{1}{4\epsilon^2}, \pm \frac{1}{2\epsilon^2}\} \end{aligned} \quad (2.6)$$

Thus for the 2D matting problem, we search for solutions of $\alpha \in H^1(\Omega)$, additionally constrained by the above properties (2.4)-(2.6). Henceforth we will refer to the above properties as *transition-function* conditions.

3. Well Posedness. In this section, we wish to compare the stability of solutions computed using our α -regularization, with the usual choice of regularizing $\nabla\alpha$ used in many matting algorithms. We will look at illustrative examples for the 1D (Fig 2.1) and 2D (Fig 2.2) cases for the inverse problem corresponding to the forward model $I = \alpha F$. Here we have ignored the background component B in (2.2), and only look at recovering (F, α) from a given I .

3.1. 1D case. In Fig 2.1 (a) is shown a 1D image I , and the corresponding ground truth α (red curve in (b)) and F (blue curve in (b)). As described previously, α is a transition function, i.e $\alpha(x) = \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} u(y)dy$ for a binary valued $u(x)$ and a window size ϵ . The values of F were chosen from a known gaussian distribution.

Given such an I , we consider the inverse problem of recovering α and F , using a gaussian prior for F , and for two different regularization choices for α (results shown in (c),(d)). In (c), we search for $\alpha \in BV(\Omega)$, and in (d), α is searched for in the space

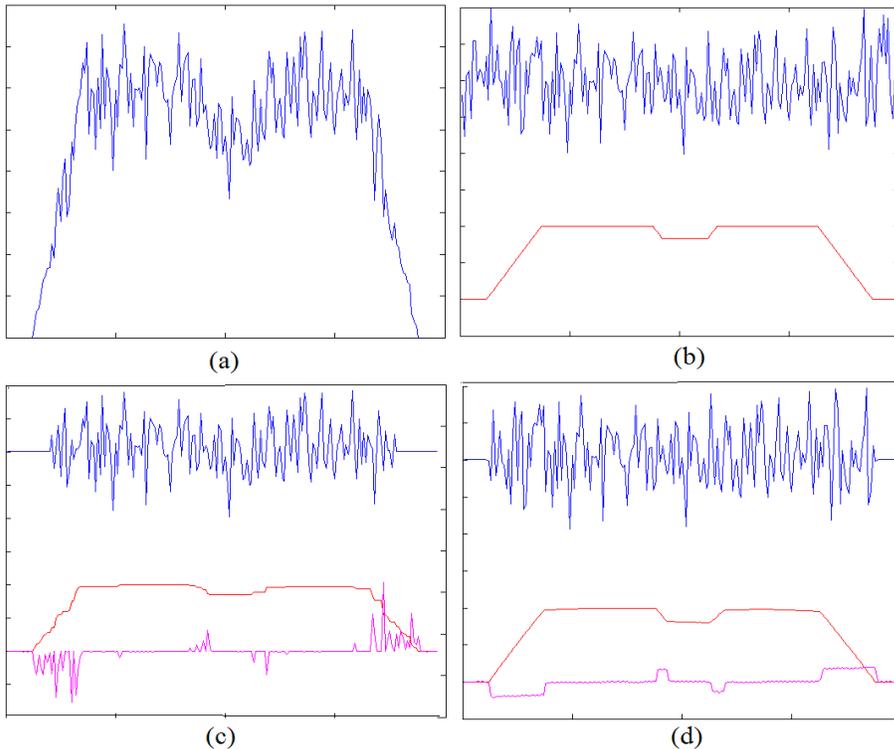


FIG. 2.1. 1D matting example; In (b)-(d): F (Blue), α (Red), α' (Magenta)

of transition functions, i.e. $\alpha \in BV(\Omega)$ with constraints (2.3). In (c) and (d), the blue, red and magenta plots are the graphs of F , α , and α' respectively.

By restricting the search space of α only to transition functions, we obtain stable convergence to solution (d) that is close to the ground truth, even with rough initial guesses. However use of the search space such as $BV(\Omega)$ for α is not sufficient to guarantee uniqueness, and leads to stability issues (e.g, as in Fig 2.1(c), even a very close initial guess for α and F can give an incorrect solution!).

3.2. 2D Case. In Fig 2.2 (a) is shown the given image I . The corresponding ground truth α and F are seen in first row (b and c). In (a), the cyan and red curves overlaid on the image is the trimap initial guess. The interior of the cyan curve is the region $\{\alpha = 1\}$, and likewise the exterior of the red curve gives the region $\{\alpha = 0\}$. Using a trimap, one can easily generate initial guesses for α (shown in (b)) and F (shown in (c)). The resulting α (i.e. the matte) is shown in (d).

Fig 2.2 (first row), shows the results obtained by using the search space $BV(\Omega)$ for α . In the second and third rows, we show the results for $\alpha \in BV(\Omega)$ additionally constrained by (2.4) - (2.6). In both cases, F has the search space, $BV(\Omega)$. The first row demonstrates the instability of a model that uses $\alpha \in BV(\Omega)$ as the search space. The ground truth shown in (b) and (c) was the initial guess for α and F . In spite of such an accurate initial guess for α and F , an incorrect solution (d) is obtained. In the second and third rows, we see that use of the transition function constraints (2.4)- (2.6) for α has given robustness with respect to initial trimap guess (e.g. two different rough initial guesses ((b) and (c)) converged to the correct α value (d)).

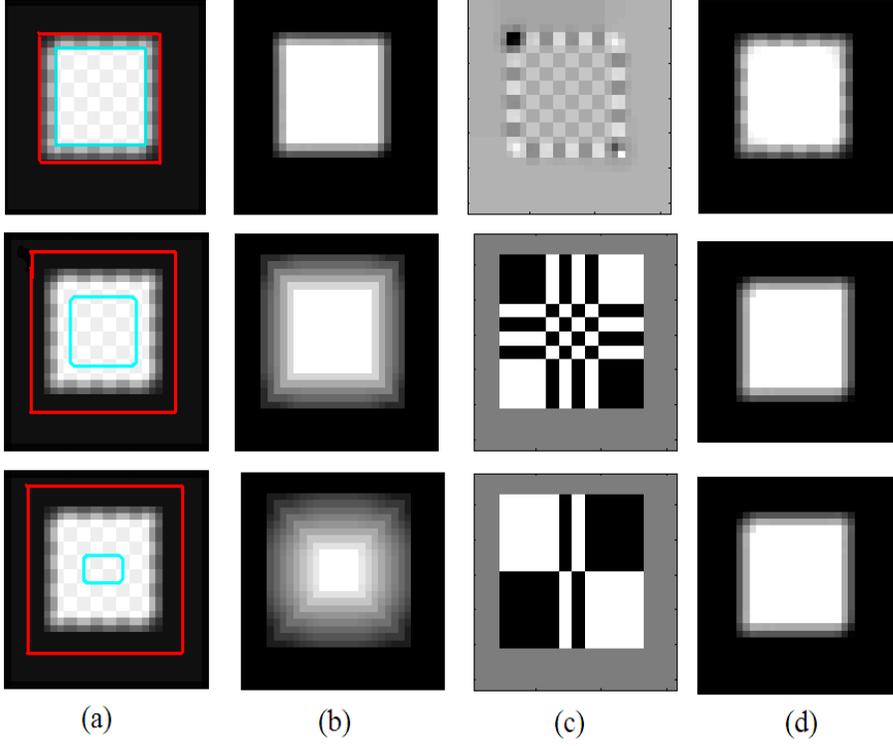


FIG. 2.2. Wellposedness: $\alpha \in BV(\Omega)$ (First Row), Our model (Second and Third Rows); (a) Image with Trimap guess ($\alpha = 1$ inside cyan curve) (b) Initial F (c) Initial α (d) α Result.

4. 2D Matting Energy. We formulate the inverse problem of recovering (F, B, α) each in $BV(\Omega)$, from I , as the following energy minimization:

$$\begin{aligned}
E(\alpha, F, B) = & \int_{\Omega} [I - (\alpha F + (1 - \alpha)B)]^2 dx + \lambda_F \int_{\Omega} |\nabla F| + \lambda_B \int_{\Omega} |\nabla B| \\
& + \beta_1 \int_{\Omega} H(|\nabla \alpha|^2 - 2c) dx + \lambda_{\alpha_1} \int_{\Omega} H(\alpha^2) H((\alpha - 1)^2) dx + \lambda_{\alpha_2} \int_{\Omega} |\nabla \alpha_{xy}| \\
& + \beta_2 \int_{\Omega} H(\alpha_{xy}^2) H((\alpha_{xy} - c)^2) H((\alpha_{xy} + c)^2) H((\alpha_{xy} - 2c)^2) H((\alpha_{xy} + 2c)^2) dx
\end{aligned} \tag{4.1}$$

Here, $c = \frac{1}{4\epsilon^2}$, and $H(t)$ is the Heaviside function. The data fidelity term follows from the matting equation. The second and third terms in the energy regularize the TV norm of F and B . The fourth term constrains $|\nabla \alpha|$ within the necessary bound $\frac{\sqrt{2}}{2\epsilon}$. The fifth term penalizes the area of the region $\alpha \notin \{0, 1\}$, thus is minimized for an $\alpha \in \{0, 1\}$. The combination of the fourth and fifth terms drives the transition region $D = \{x \in \Omega | 0 < \alpha(x) < 1\}$ to be close to the required width of 2ϵ (necessary conditions (2.4) and (2.5)). Finally, the sixth term constrains the TV norm of α_{xy} , and the last term restricts the values of α_{xy} to lie within $\{0, \pm \frac{1}{4\epsilon^2}, \pm \frac{1}{2\epsilon^2}\}$ (necessary condition (2.6)). $\lambda_F, \lambda_B, \lambda_{\alpha_1}, \lambda_{\alpha_2}$ are parameters that balance the respective terms,

and β_1, β_2 are Lagrange multipliers for the constraints. λ_F and λ_B are parameters for the TV terms and relate to the scale of objects within the foreground and background regions.

To minimize (4.1), we start with an initial guess (α_0, F_0, B_0) for the unknowns and iteratively refine the guess using steepest descent. The initial guess here is analogous to the trimap initialization required by many matting algorithms. However, our model gives robustness to initial guess due to the transition-function conditions used in (4.1), as demonstrated in the experiments section. In current matting algorithms, there are three main steps which we have combined into the above energy:

- *Trimap refinement*: A prior segmentation step is added to refine the approximate user defined trimap to conform to the actual α -transition region. In our energy (4.1), this step is accomplished through the fourth and fifth terms.
- *Extrapolating F and B* : The foreground and background intensities, F and B have to be extrapolated (i.e inpainted) into the transition region. Our inpainting technique is TV based that is successful in inpainting geometric features from the known regions.
- *Solving for α* : α is solved for in the transition region using the matting equation subject to suitable priors. We constrain the smoothness of α additionally by the fifth and sixth terms, that comes from the image formation model discussed in the previous section. Such a reduced space is essential for the stability and local uniqueness of the computed solutions.

5. Experimental Results. In this section, we present results on some synthetic data (Fig. 5.1, Fig. 5.2, Fig. 5.3) and on real images (Fig.5.4, Fig.5.5, Fig.5.6). In the synthetic examples Fig. 5.1 and Fig. 5.2, the foreground and background intensities are from gaussian distributions. In Fig. 5.1, the use of α regularization shown in (4.1) resulted in an α (shown in (d)) very close to the ground truth matte (b), although only a rough initial guess was used. Whereas, in (c) we see that a typical regularization of $\nabla\alpha$ such as the $TV(\alpha)$ is not sufficient to avoid local minima (even with a good initial guess!), and could result in a noisy matte. In another example (Fig. 5.1), the model shows robustness to the choice of F and B (in this case with large variances), and recovers the matte accurately. In image Fig. 5.3 (a), the background has a sharp discontinuity. A rough trimap guess (Definitely foreground - interior of Cyan contour, Definitely background - exterior of Red contour) is used. The initialization for B is shown in (b). We see the result for α (in (c)) and B (in (d)). Proper inpainting of B utilizing the geometry in the known region and driven by the image intensity, as in our energy (4.1) is crucial for a good matte result. Nearest neighbour based inpainting techniques (frequently used in matting algorithms) cannot give such a result without a careful choice for the trimap.

For real images (Fig.5.4, Fig.5.5, Fig.5.6)(images shown in (a)), we start with a trimap guess (overlaid in (a)), use the trimap to generate initial estimates for F (shown in (b)), B (shown in (c)). The minimizing F , B and the matte α of the energy (4.1) are shown in (d)-(f). The extracted object (i.e. αF) is shown in (h).

In Fig.5.5 (i), we see an artifact in the extracted object(highlighted by cyan colored square) when using an iterative algorithm such as Poisson matting that regularizes $\nabla\alpha$ and uses nearest neighbour inpainting.

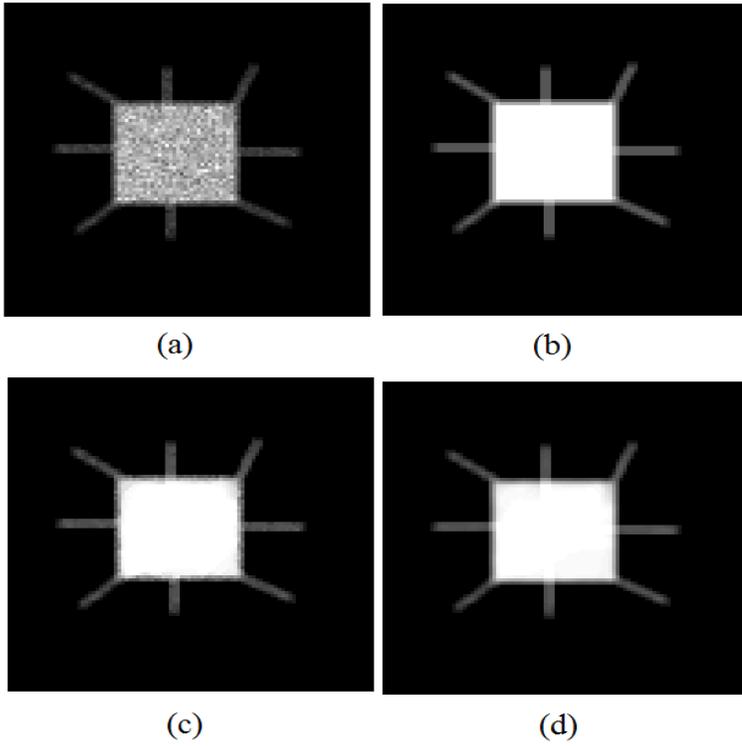


FIG. 5.1. (a) Image (b) α regularity as in (4.1) (c) TV regularity for α (d) ground truth

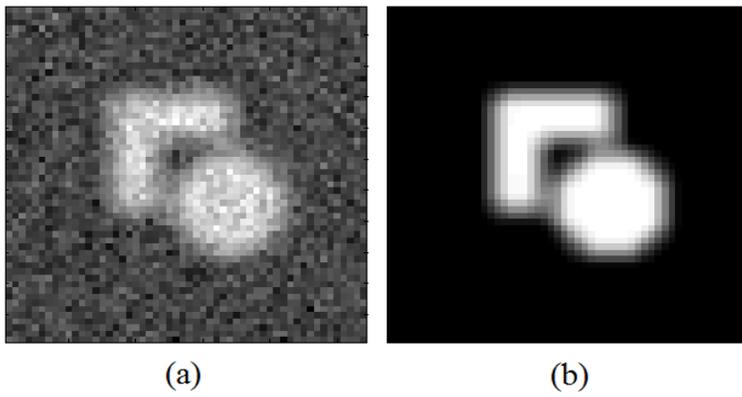


FIG. 5.2. (a) Image with noisy F and B (b) α result

REFERENCES

- [1] Nicholas Apostoloff and Andrew Fitzgibbon. Bayesian video matting using learnt image priors. In *Proc. Conference on Computer Vision and Pattern Recognition*, volume 1, pages 407–414, 2004.
- [2] A. Berman, P. Vlahos, and A Dadourian. Comprehensive method for removing from an image the background surrounding a selected object. *U.S. Patent 6,134,345*, 2000.
- [3] Yuri Boykov and Marie Pierre Jolly. Interactive graph cuts for optimal boundary and region

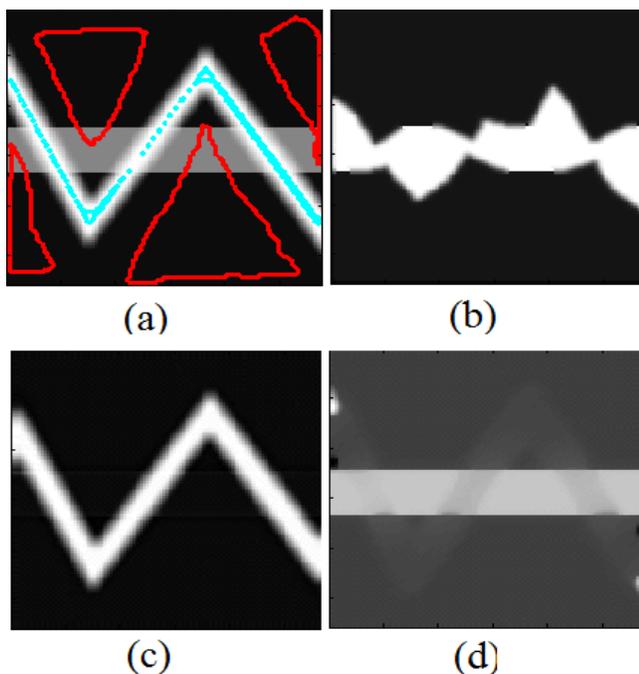


FIG. 5.3. (a) Image with Trimap guess (b) Initial B (c) α result (d) B result

- segmentation of objects in n-d images. In *Proc. International Conference on Computer Vision*, pages 105–112, 2001.
- [4] Yung-Yu Chuang, Brian Curless, David H. Salesin, and Richard Szeliski. A bayesian approach to digital matting. In *Proc. Conference on Computer Vision and Pattern Recognition*, volume 2, pages 264–271, 2001.
 - [5] Raymond Fielding. *The Technique of Special-effects Cinematography*. Hastings House, 1965.
 - [6] Yu Guan, Wei Chen, Xiao Liang, Zi’ang Ding, and Qunsheng Peng. Easy matting - a stroke based approach for continuous image matting. In *Proc. EUROGRAPH*, pages 567–576, 2006.
 - [7] Samuel W. Hasinoffa, Sing Bing Kang, and Richard Szeliski. Boundary matting for view synthesis. *Computer Vision and Image Understanding*, 103(1):22–32, 2006.
 - [8] P. Hillman, J. Hannah, and D. Renshaw. Alpha channel estimation in high resolution images and image sequences. In *Proc. Conference on Computer Vision and Pattern Recognition*, volume 1, pages 1063–1068, 2001.
 - [9] K.Ni, S.R. Thiruvankadam, and T.F. Chan. Matting through variational inpainting. In *SIP*, pages 173–179, 2007.
 - [10] Anat Levin, Dani Lischinski, and Yair Weiss. A closed form solution to natural image matting. In *Proc. Conference on Computer Vision and Pattern Recognition*, volume 1, pages 61–68, 2006.
 - [11] Yin Li, Jian Sun, Chi-Keung Tang, and Heung-Yeung Shum. Lazy snapping. *ACM Transactions on Graphics*, 23(3):303–308, 2004.
 - [12] Y. Mishima. Soft edge chroma-key generation based upon hexoctahedral color space. *U.S. Patent 5,355,174*, 1993.
 - [13] Thomas Porter and Tom Duff. Compositing digital images. In *Proc. SIGGRAPH*, pages 253–259, 1984.
 - [14] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. “grabcut”: interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3):309–314, 2004.
 - [15] Mark A. Ruzon and Carlo Tomasi. Alpha estimation in natural images. In *Proc. Conference on Computer Vision and Pattern Recognition*, volume 1, pages 18–25, 2000.
 - [16] Alvy Ray Smith and James F. Blinn. Blue screen matting. In *Proc. SIGGRAPH*, pages 259–268, 1996.

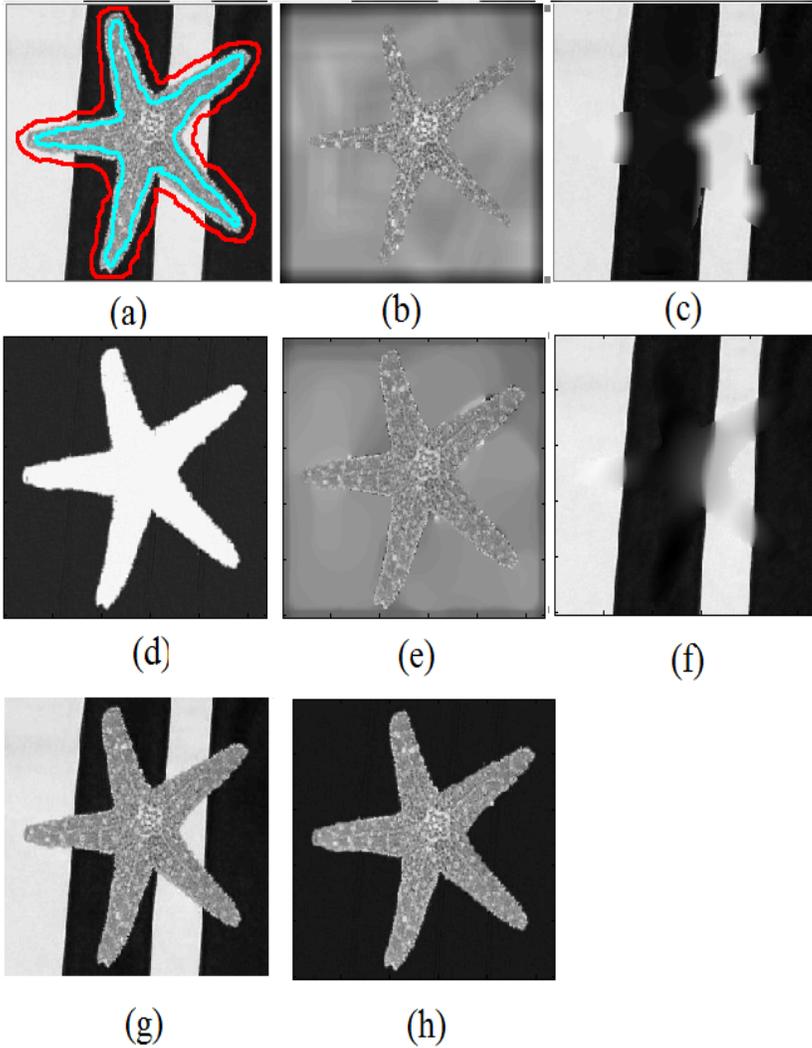


FIG. 5.4. (a) Starfish image with Trimap guess (b),(c) Initial Guesses for, F and B (d)-(f) Result for: α , F , and B (g) Original Image (h) Extracted object αF

- [17] Jian Sun, Jiaya Jia, Chi-Keung Tang, and Heung-Yeung Shum. Poisson matting. *ACM Transactions on Graphics*, 23(3):315–321, 2004.
- [18] Jian Sun, Yin Li, Sing Bing Kang, and Heung-Yeung Shum. Flash matting. *ACM Transactions on Graphics*, 25(3):772–778, 2006.
- [19] J. Wang and M.F. Cohen. An iterative optimization approach for unified image segmentation and matting. In *Proc. International Conference on Computer Vision*, pages 936–943, 2005.
- [20] Y. Wexler, A. Fitzgibbon, and A. Zisserman. Bayesian estimation of layers from multiple images. In *Proc. European Conference on Computer Vision*, pages 49–61, 2002.
- [21] J. Xiao and M. Shah. Accurate motion layer segmentation and matting. In *Proc. Conference on Computer Vision and Pattern Recognition*, volume 2, pages 698–703, 2005.

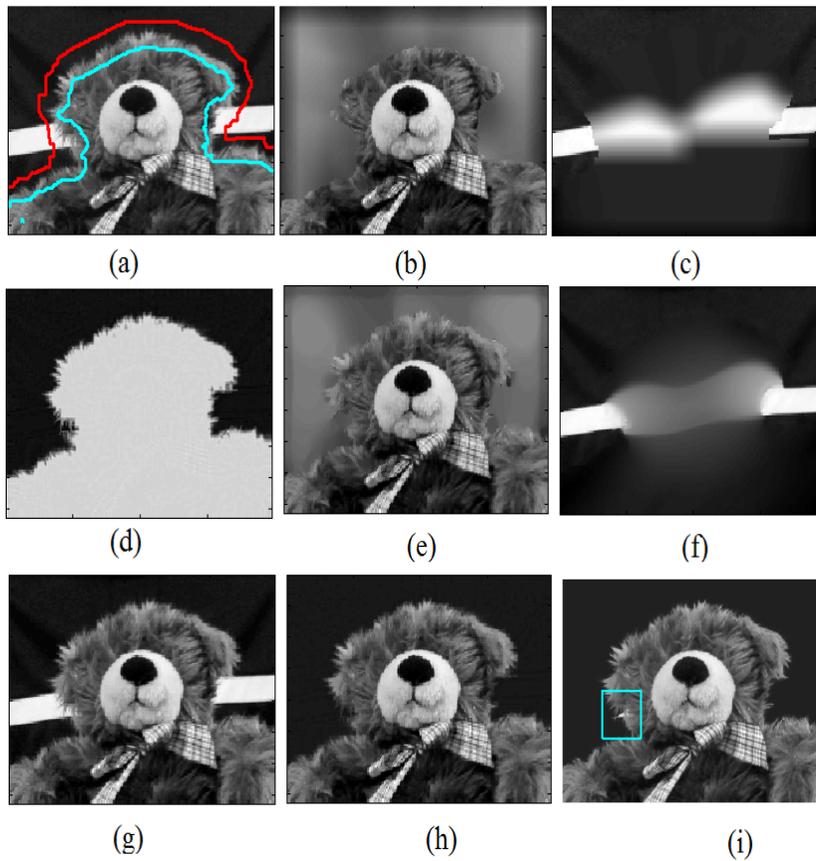


FIG. 5.5. (a) Bear image with Trimap guess (b),(c) Initial Guesses for, F and B (d)-(f) Result for: α , F , and B (g) Original Image (h) Extracted object αF (i) Result for an iterative algorithm similar to Poisson matting

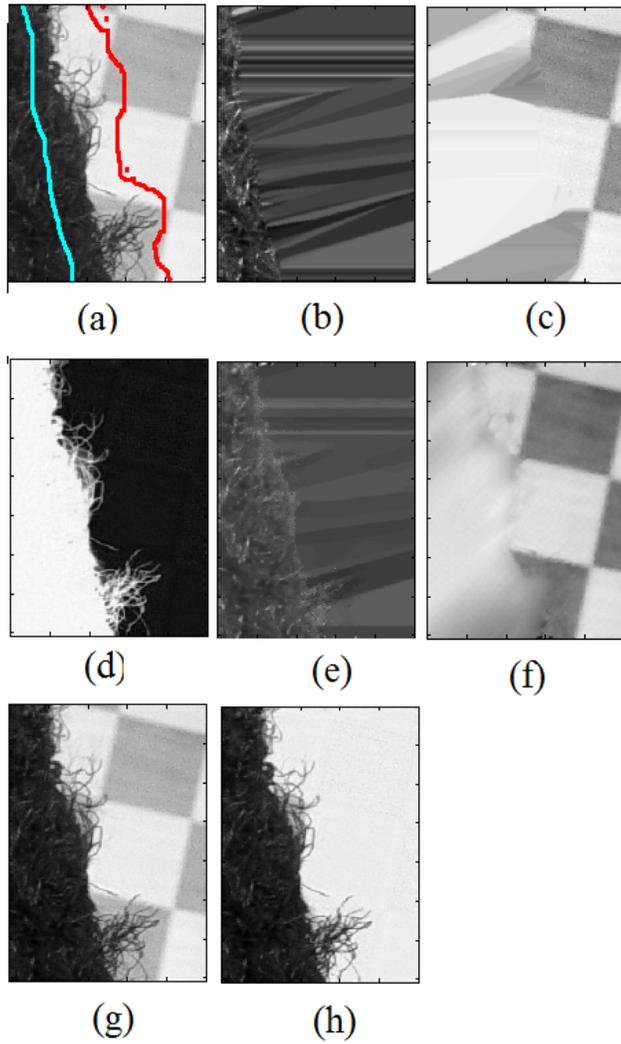


FIG. 5.6. (a) Scarf image with Trimap guess (b),(c) Initial Guesses for, F and B (d)-(f) Result for: α , F , and B (g) Original Image (h) Extracted object αF