

TVSeg - Interactive Total Variation Based Image Segmentation

Markus Unger¹, Thomas Pock^{1,2}, Werner Trobin¹,
Daniel Cremers², Horst Bischof¹

¹Inst. for Computer Graphics & Vision, Graz University of Technology

²Department of Computer Science, University of Bonn

Abstract

Interactive object extraction is an important part in any image editing software. We present a two step segmentation algorithm that first obtains a binary segmentation and then applies matting on the border regions to obtain a smooth alpha channel. The proposed segmentation algorithm is based on the minimization of the Geodesic Active Contour energy. A fast Total Variation minimization algorithm is used to find the globally optimal solution. We show how user interaction can be incorporated and outline an efficient way to exploit color information. A novel matting approach, based on energy minimization, is presented. Experimental evaluations are discussed, and the algorithm is compared to state of the art object extraction algorithms. The GPU based binaries are available online.

1 Introduction

In this paper, we address the problem of interactive image segmentation of arbitrary still images. The process of separating an image into foreground and background regions is often also called object extraction. Ideally, such an extraction is not done using a hard, binary labeling, but by finding a smooth alpha channel, known as image matting. This way, also transparent regions and fine details such as hair can be sufficiently extracted. As already pointed out in [14], pure alpha-matting approaches work fine if there are discriminative color distributions for foreground and background, but they often fail in camouflage. We therefore decided to develop a two step algorithm that first finds a binary labeling into foreground and background. This is done using fast *Total Variation* minimization of the *Geodesic Active Contour* energy with local constraints. If desired, a second step performs alpha-matting along the border of the binary segmentation.

The remainder of this paper is organized as follows: First we discuss related work, and outline the fundamental ideas that led to our algorithm in Section 2. In Section 3, we discuss the algorithm, show how local constraints can be incorporated and describe an efficient way to utilize color information. Section 4 presents results on different images and provides a comparison to state of the art image extraction algorithms. Finally, Section 5 gives a short conclusion and discusses possible directions for future investigations.

2 Related Work

2.1 Geodesic Active Contours

Caselles et al. introduced the Geodesic Active Contour (*GAC*) model in [5], as an enhanced version of the snake model of Kass et al. [10]. The *GAC* model is defined as the following variational problem:

$$\min_C \left\{ E_{GAC}(C) = \int_0^{|C|} g(|\nabla I(C(s))|) dl \right\}, \quad (1)$$

where $|C|$ is the Euclidean length of the curve C and dl the Euclidean element of length. The function $g \in (0, 1]$ is an edge detection function with values close to 0 at strong edges in the image I . Thus, the *GAC* model integrates the Euclidean element of length dl weighted by a term depending on the edge information of the image. By the definition of E_{GAC} we can see that the trivial solution $C = \emptyset$ is always a global minimizer. Hence the *GAC* model yields meaningful results only in combination with additional constraints.

To find the globally optimal solution of (1), graph based approaches are commonly used. They rely on the partitioning of a graph that is built based on the image. Boykov et al. [3] used a minimum cut algorithm to separate a graph around foreground and background terminals. With the right choice of the edge weights, and for successively finer grids, the Euclidean length of the contour can be approximated. It is well-known that the quality of this approximation highly depends on the order of connectivity of the underlying graph, and can lead to systematic metrication (discretization) errors. Other graph based approaches were e.g. taken by Grady with the random walker algorithm [9] and its derivatives [16]. Appleton and Talbot presented an approach, that minimizes the *GAC* energy using continuous maximal flows [1]. This approach does not suffer from any discretization errors.

A different approach, which uses the *weighted Total Variation* was introduced by Bresson et al. in [4]. The *weighted TV* is defined as

$$TV_g(u) = \int_{\Omega} g(x) |\nabla u| d\Omega. \quad (2)$$

Bresson et al. showed that if u is a characteristic function 1_C , (2) is equivalent to E_{GAC} in (1). Note that the characteristic function 1_C is a closed set in the image domain Ω and C stands for its boundary. If u is allowed to vary continuously between $[0, 1]$, (2) becomes a convex functional, meaning that one can compute the global minimizer of it. To obtain the final segmentation, a level set of u is selected (we generally chose $u = 0.5$). With the *weighted TV* formulation no metrication errors are introduced.

As already mentioned above, the pure *GAC* energy has to be constrained in order to obtain meaningful results. In [11], Leung and Osher unified denoising, segmentation and inpainting. Their idea was to use (2) together with a spatially varying L^1 data fidelity term. In [18, 19], a way to incorporate various local constraints was shown. It turns out that fast *TV* minimization leads to a globally optimal, fast and highly interactive segmentation algorithm. The *weighted Total Variation* based algorithm suggested in [18] already showed promising results on different medical grayscale 2D and 3D datasets.

2.2 Image Segmentation Frameworks

The *GrabCut* algorithm [14] was suggested by Rother et al. in 2004. It is a two step algorithm that first finds a binary labeling into foreground and background and then performs matting in the border regions. In the first step, the binary segmentation is obtained by minimizing a regularized energy. Foreground and background are represented by separate *Gaussian Mixture Models (GMM)*. By drawing a rectangle, everything outside is set as background. Iteratively, the *GMM* is assigned to the pixels, then updated and finally the segmentation is estimated using a minimum cut algorithm. The algorithm is guaranteed to converge at least to a local minimum [14]. Additionally, hard foreground and background labels can be incorporated during the segmentation process. In the matting step, a smooth transition is obtained by minimizing an energy consisting of a data term that fits a soft step function and a regularization term.

Also *SIOX* [8] was designed for interactive object extraction. The idea is to use a trimap indicating foreground, background and unspecified regions. For the known regions, a color signature is calculated and clustering is performed. After that, a weighted nearest neighbor search is used to classify the pixels. In a post-processing step artifacts are removed and small regions are purged.

Other interactive image segmentation algorithms were e.g. developed by Cremers et al. in [7]. Here a probabilistic level set formulation is used. In [2], Bai et al. proposed a geodesic based image segmentation and matting framework.

2.3 Image Matting

Image matting is based on the assumption that the input image I is a composite of the foreground image I_F and the background image I_B . One assumes that the color of a pixel x is a linear combination of the foreground and background color:

$$I(x) = \alpha(x)I_F(x) + (1 - \alpha(x))I_B(x), \quad (3)$$

with α being the pixel's foreground opacity. Note that this problem is severely underconstrained, as only I is known. To solve this matting problem, assumptions on the image have to be made.

The Bayesian matting approach, proposed by Chuang et al. in [6], models foreground and background by spatially varying Gaussian models, and α is obtained using maximum-likelihood estimation. In [12], Levin et al. presented a closed form solution to solve the matting problem with promising results. Using local smoothness assumptions they derived a quadratic cost function for α . By solving a sparse linear system, the global optimum of this cost function is obtained. Constraints can be easily incorporated, and interaction is done using different brushes. To be computationally feasible for large images, a multigrid solver is used, resulting in degradation of small structures. In [17], Thiruvankadam et al. developed a matting approach based on a variational PDE formulation. An energy minimization problem in three variables (α , I_F and I_B) has to be solved. They used *TV* regularization on all three variables. As initialization, a trimap is provided to the algorithm. The algorithm has the disadvantage that six parameters are necessary to balance the terms in the energy.

3 Proposed Algorithm: TVSeg

3.1 Step 1: Binary Segmentation

As in [19], we minimize the following variational image segmentation model:

$$\min_{u \in [0,1]} \left\{ E_{Seg} = \int_{\Omega} g(x) |\nabla u| d\Omega + \int_{\Omega} \lambda(x) |u - f| d\Omega \right\}. \quad (4)$$

Here, the first term of the energy is the *weighted TV* of u as defined in (2), which minimizes the *GAC* energy. The second term is used to incorporate constraints into the energy functional. The variable $f \in [0, 1]$ is provided by the user and indicates foreground ($f = 1$) and background ($f = 0$) seed regions. The spatially varying parameter $\lambda(x)$ is responsible for the interpretation of the information contained in f . Three different cases for $\lambda(x)$ can be discriminated:

- $\lambda(x) = 0$: Here, the right hand term of E_{seg} vanishes, and the information in f is not considered by the algorithm. Thus the pure *GAC* energy is computed.
- $\lambda(x) \rightarrow \infty$: The information in f is considered as hard constraint. The *weighted TV* term is ignored, and the right hand term of E_{seg} forces $u = f$.
- $0 < \lambda(x) < \infty$: While still the *GAC* energy is minimized, the constraints in f get approximated. Therefore we refer to this case as weak constraints. The smaller the value of $\lambda(x)$ the stronger the approximation.

The incorporation of different constraints enables the user to interact with the algorithm. For this purpose, different brushes are provided:

- Hard foreground and background brush: These brushes set f to the according value, and additionally set $\lambda(x) = \infty$. We refer to these constraints as hard constraints, as they force u to a certain value.
- Sample foreground and background brush: Using this brush we are able to exploit the color information. The drawn sampling regions are used to build 3D color histograms based on the RGB color space. These histograms characterize foreground $F(r, g, b)$ and background $B(r, g, b)$. This allows us to compare the probability that a pixel x was seen in the foreground $F(r(x), g(x), b(x))$, to the probability that the pixel was seen in the background $B(r(x), g(x), b(x))$. We thus get the histogram match $\Delta_H = F(r(x), g(x), b(x)) - B(r(x), g(x), b(x))$, and for each pixel x we set:

$$\begin{aligned} f(x) &= 1 & \text{if } \Delta_H > \theta_h \\ f(x) &= 0 & \text{if } \Delta_H < -\theta_h \\ \lambda(x) &= 0 & \text{if } |\Delta_H| \leq \theta_h \end{aligned} \quad (5)$$

The third case ensures that if the probabilities are very similar (measure specified by a threshold θ_h) the information in f is ignored by setting $\lambda(x) = 0$. In all other cases the user can specify λ globally and thus determine the amount of approximation that is applied to the obtained regions.

- Draw and erase edges: By modifying the edge image, the user has the control to directly influence the weight function used to minimize the *GAC* energy.

In [19], a fast iterative scheme to minimize the energy defined in (4) was presented. This is done using a projected gradient descend algorithm based on a dual formulation of E_{seg} . Here, we use the same numerical implementation to obtain the globally optimal solution.

The edge detection function g was computed as $g(I) = \exp(-a|\nabla I|^b)$, with $a = 10$ and $b = 0.55$. To improve the results, we applied *ROF* denoising [15] on I before edge calculation. This guarantees that image noise does not interfere with the edge image, and the *GAC* energy is computed only on significant edges.

3.2 Step 2: Matting

The binary segmentation u_0 obtained in the first step is used as input for the matting step, by calculating a region around the contour C where matting will be performed. This is achieved using an Euclidean signed distance transformation $sd(x, C)$, that is positive inside the region and negative outside. Thus we obtain an inner region $D_F := \{x \in \Omega | sd(x, C) > \theta\}$ where α is fixed as foreground, and a region indicating fixed background $D_B := \{x \in \Omega | sd(x, C) < -\theta\}$. θ is a threshold specifying the border width. As a consequence, the region where matting is performed is defined as $D_M := \{\Omega \setminus (D_F \cup D_B)\}$. In addition to the automatic matting regions, brushes are provided to interactively add or delete matting regions.

For matting, we propose to minimize the following energy:

$$\min_{\alpha, I_F, I_B} \left\{ E_{Mat} = \sum_c \left[\frac{1}{2} \int_{\Omega \setminus D_F} |\nabla I_{F,c}|^2 + \frac{1}{2} \int_{\Omega \setminus D_B} |\nabla I_{B,c}|^2 + \eta \int_{\Omega} |\nabla \alpha|^2 + \frac{\beta}{2} \int_{D_M} (I_c - \alpha I_{F,c} - (1 - \alpha) I_{B,c})^2 + \frac{\gamma}{2} \int_{\Omega} (\alpha - u_0)^2 \right] \right\}. \quad (6)$$

Here the image I_F is the smooth, approximated foreground, and I_B contains the smoothed background. Both images are determined automatically by the minimization scheme. See Figure 3 for an exemplary result after convergence. The fourth term ensures that the matting model as defined in (3) is solved. With the last term, the matting result is ensured to be close to the binary labeling u_0 . The third term of E_{Mat} performs regularization on the α channel. The free parameters η , β , and γ are used to balance the single terms.

For minimization, we derive the associated Euler-Lagrange equations and apply a standard semi-implicit fixed point scheme. This results in an iterative three-step algorithm, where alternated minimization is done for I_F , I_B and α .

4 Results

The proposed algorithm was implemented on the *GPU*. The application is publicly available online¹. We tested the algorithm on a wide range of images. Subsequently some results are presented and discussed. For all the images presented below we chose $\lambda = 0.1$. As this algorithm is designed as an interactive tool, no exact speed evaluations were done. But note that the algorithm always reacts immediately on new user input. For the images presented here, convergence time is marginal compared to the interaction time.

¹<http://www.gpu4vision.org>

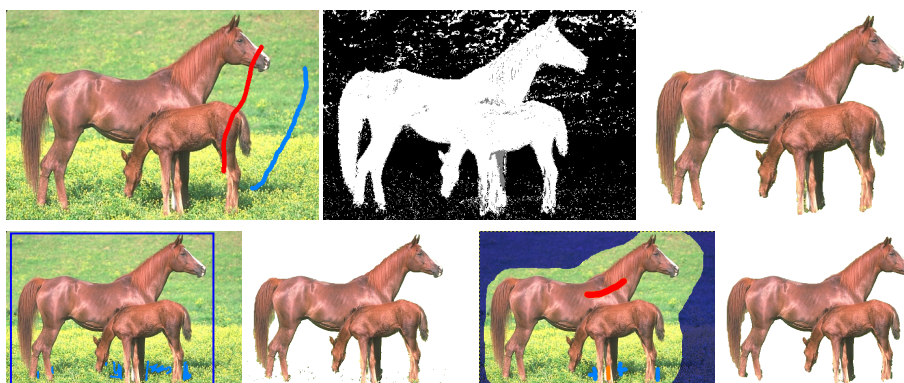


Figure 1: Top row: Segmentation process of the *Horse* image, using our algorithm. On the left, the input image is shown together with the sampling regions, foreground as red and light blue as background. The middle image shows the variable f , and the right image shows the extracted image. Bottom row: The two images on the left show the interaction needed by the *GrabCut* implementation, and the extraction result. The right images, show the segmentation process using the *SIOX* implementation.

4.1 Segmentation Results

The images shown in this section were obtained solely by the binary segmentation step. Matting could sometimes improve the result, by removing small speckles that may occur at strong edges. The top row of Figure 1 shows how the proposed algorithm exploits color information by using the sampling brushes. The middle image shows f , and is obtained by histogram matching as described in Section 3.1. White areas of f are foreground seeds, black areas are background seeds, and in the gray area the pure *GAC* energy is minimized. One can also see that the sample regions, incorporated by the provided brushes, do not need to be very exact to obtain the desired segmentation result.

We also compared our implementation to a publicly available *GrabCut* [14] implementation², and the *SIOX* [8] implementation integrated in *Gimp 2.4.2*. The bottom row of Figure 1 shows these two algorithms applied to the *Horse* image [13]. Note that both algorithms deliver a binary labeling. In this example the segmentation results are almost identical for the three algorithms. Only the *GrabCut* algorithm produces a lot of speckles, that could be easily removed. Also note that our algorithm needs slightly less user interaction.

Figure 4 shows more images segmented with the three algorithms. In the first row the *Soldier* image, taken from [13], was segmented. We used sampling as well as hard constraint brushes (light red for foreground, and dark blue for background). For the helmet, an additional edge (light green) was introduced. Overall, all three algorithms deliver similar results. Note that due to the high quality of the edge image g in the *weighted TV* term, our algorithm produces a very exact segmentation border. The degree of interaction for the *GrabCut* algorithm and our approach is approximately the same. Only the *SIOX* algorithm needs slightly more user interaction. When looking at the other rows of

²<https://mywebspaces.wisc.edu/pwang6/personal/> (22 Apr. 2008)



Figure 2: Segmentation of the *Teddy* image taken from [12]. From left to right: The original image with overlaid sample regions. The final alpha map using only the segmentation step. Alpha map using matting around the border. Sample regions used by Levin et al. [12], and the according alpha map.

Figure 4, it turns out that our algorithm generally needs the least user interaction. Furthermore, our algorithm is highly interactive and already reacts to the user input while a new constraint is drawn. This results in a more efficient interaction. When the input image is a grayscale image, as shown in the last row of Figure 4, the proposed algorithm performs significantly better than the other algorithms. In this case our algorithm relies solely on the *GAC* energy and the hard constraints. It turns out that the performance of the *SIOX* algorithm breaks down if no color information is available. While the *GrabCut* algorithm still delivers appropriate results, significantly more interaction is needed.

4.2 Matting Results

As discussed earlier, a binary labeling approach is not always desirable, especially if fine structures like hair, or transparent regions should be segmented. In Figure 2, matting results using our implementation are shown on the *Teddy* image where hair forms the segmentation border. One can clearly see that the binary labeling is not sufficient if the results are meant to be used for further image compositing. We compared the results of our implementation to the matting approach by Levin et al. [12], and our algorithm delivers competitive results. Unfortunately, parametrization of our algorithm has to be chosen for each image individually to obtain optimal results.

A second matting example is shown in Figure 3. After the binary segmentation is performed in the first step, the matting regions are automatically determined, and the matting energy (6) is minimized. The matting result needed no further refinement. As the right image of Figure 3 shows, the results are well suited for image compositing. Figure 3 also depicts the images I_F and I_B . Where I_F is the smooth foreground approximation, and I_B the smooth background approximation.

5 Conclusion

As the experimental results showed, the proposed segmentation framework is able to perform exact object extraction on a wide range of images. This is done interactively using a minimum amount of user input. The strength of the proposed segmentation framework is the utilization of different features. By using the sampling brushes and histogram matching, we showed how color information can be efficiently exploited. The contour is smoothed using robust edge information contained in the image. Even if no color



Figure 3: Segmentation of the *Rabbit* image taken from [12]. From left to right: Original image with constraints and binary segmentation border overlaid in red. Automatic determination of matting regions. The resulting α -map. The image I_F and I_B . A composition of the extracted image with a new background.

information is available, our algorithm delivers fast and exact results by relying on hard constraints. The comparison of the proposed framework to state of the art algorithms emphasized the benefits of our approach. We also showed that the proposed matting approach is able to yield competitive results. This enables the algorithm to segment fine details and transparent regions.

A great advantage of these methods are their great parallelization potential. As current computer hardware gets more and more parallelized (see the recent development of GPUs), these methods are perfectly suited for future devices.

Future research will include coarse to fine approaches for solving (4), as large images will clearly benefit from such an approach.

6 Acknowledgement

This work was partly supported by the Austrian Research Promotion Agency (FFG) within the VM-GPU Project No. 813396, and the Hausdorff Center for Mathematics. Furthermore we like to thank NVidia for their hardware support.

References

- [1] B. Appleton and H. Talbot. Globally minimal surfaces by continuous maximal flows. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(1):106–118, 2006.
- [2] X. Bai and G. Sapiro. A geodesic framework for fast interactive image and video segmentation and matting. In *Proc. of ICCV 2007*, pages 1–8, 2007.
- [3] Y. Boykov and M.-P. Jolly. Interactive organ segmentation using graph cuts. In *Proc. of MICCAI 2000, LNCS 1935*, pages 276–286. Springer, 2000.
- [4] X. Bresson, S. Esedoglu, P. Vandergheynst, J. Thiran, and S. Osher. Global minimizers of the active contour/snake model. In *Free Boundary Problems (FBP): Theory and Applications*, 2005.
- [5] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *Intl. J. of Computer Vision*, 22(1):61–79, 1997.

- [6] Yung-Yu Chuang, B. Curless, D.H. Salesin, and R. Szeliski. A bayesian approach to digital matting. *Proc. of CVPR 2001*, 2:II-264-II-271, 2001.
- [7] D. Cremers, O. Fluck, M. Rousson, and S. Aharon. A probabilistic level set formulation for interactive organ segmentation. In *Proc. of SPIE Medical Imaging*, San Diego, USA, February 2007.
- [8] G. Friedland, K. Jantz, and R. Rojas. SIOX: Simple interactive object extraction in still images. In *Proc. of the Seventh IEEE International Symposium on Multimedia*, pages 253–260, Washington, DC, USA, 2005.
- [9] L. Grady. Random walks for image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, Nov. 2006.
- [10] M. Kass, A.P. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Intl. J. of Computer Vision*, 1(4):321–331, 1988.
- [11] S. Leung and S. Osher. Fast global minimization of the active contour model with TV-inpainting and two-phase denoising. In *3rd IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision*, pages 149–160, 2005.
- [12] A. Levin, D. Lischinski, and Y. Weiss. A closed-form solution to natural image matting. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(2):228–242, Feb. 2008.
- [13] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. of ICCV 2001*, volume 2, pages 416–423. IEEE Computer Society, July 2001.
- [14] C. Rother, V. Kolmogorov, and A. Blake. GrabCut: interactive foreground extraction using iterated graph cuts. In *SIGGRAPH '04*, pages 309–314, New York, NY, USA, 2004. ACM.
- [15] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.
- [16] A. K. Sinop and L. Grady. A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In *Proc. of ICCV 2007*. IEEE Computer Society, Oct. 2007.
- [17] S. R. Thiruvenkadam, T. F. Chan, and B. Hong. Variational approach to natural image matting. *UCLA Computational and Applied Mathematics Reports*, December 2007.
- [18] M. Unger, T. Pock, and H. Bischof. Continuous globally optimal image segmentation with local constraints. In *Computer Vision Winter Workshop 2008*, Moravske Toplice, Slovenija, February 2008.
- [19] M. Unger, T. Pock, and H. Bischof. Interactive globally optimal image segmentation. Technical Report 08/02, Inst. for Computer Graphics and Vision, Graz University of Technology, 2008.

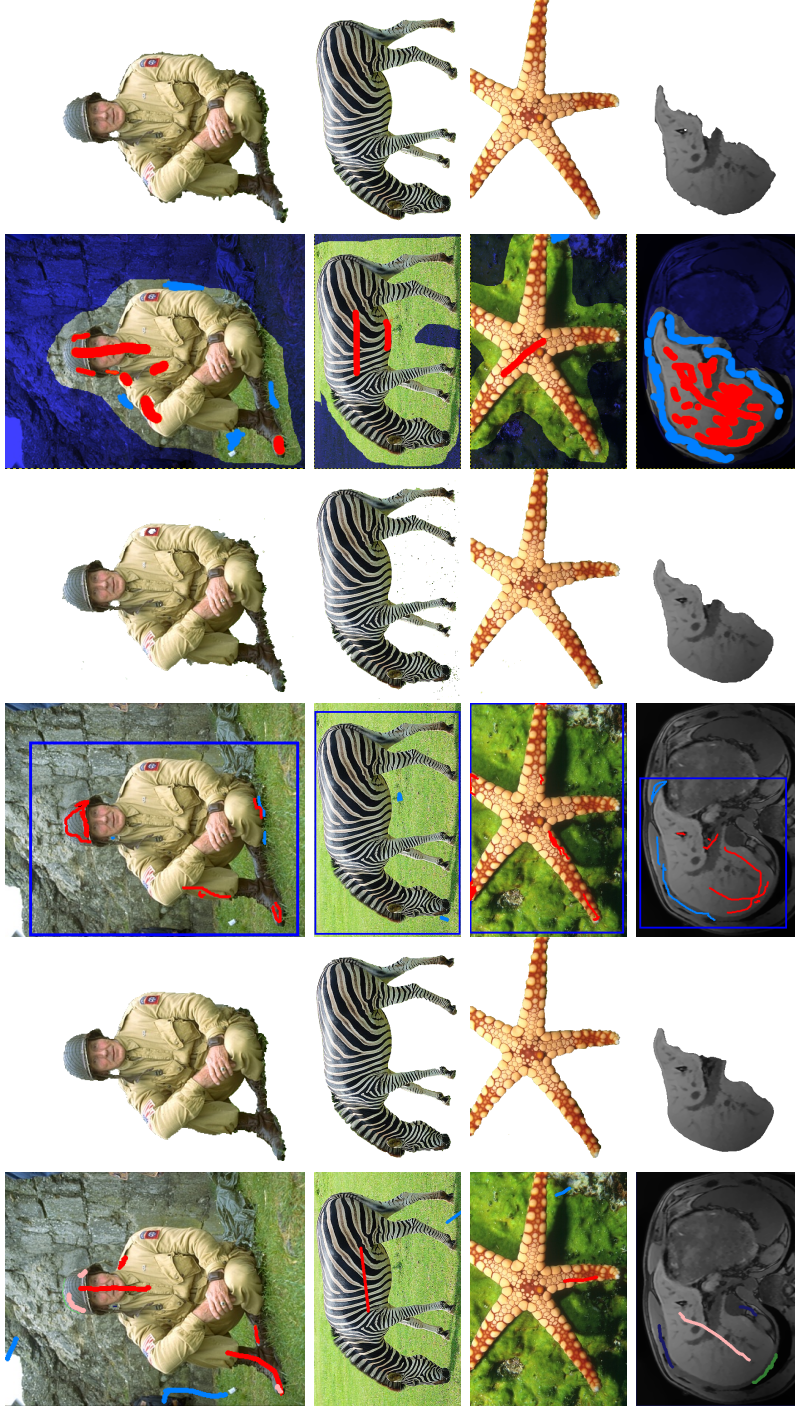


Figure 4: The proposed *TVSeg* algorithm on the left hand side, the *GrabCut* algorithm in the middle and the *SIOX* algorithm on the right hand side, applied to different images. Red areas are foreground seeds, blue areas indicate background seeds, and light green areas show modified edges. Note that *TVSeg* needs the least user input.