Optimal Solutions for Semantic Image Decomposition

Daniel Cremers

Department of Computer Science, TU Munich

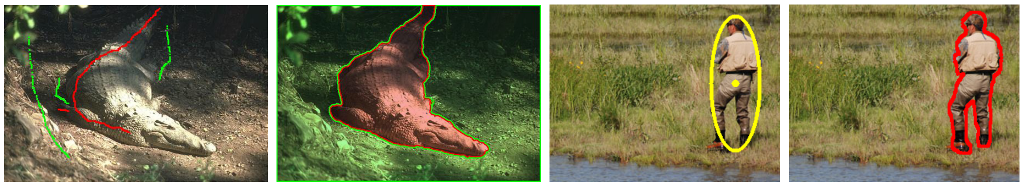
Abstract

Bridging the gap between low-level and high-level image analysis has been a central challenge in computer vision throughout the last decades. In this article I will point out a number of recent developments in low-level image analysis which open up new possibilities to bring together concepts of high-level and low-level vision. The key observation is that numerous multilabel optimization problems can nowadays be efficiently solved in a near-optimal manner, using either graph-theoretic algorithms or convex relaxation techniques. Moreover, higher-level semantic knowledge can be learned and imposed on the basis of such multilabel formulations.

Keywords: optimization, efficient algorithms, convexity, semantic labeling

1. Combining Low-level vision...

Starting in the 1980s researchers have tackled the image segmentation problem by means of energy minimization approaches [1, 12]. While early approaches were generally not convex and respective algorithms would only compute locally optimal solutions, in recent years researchers have developed algorithms to compute optimal or near optimal solutions for respective energies using either graph-theoretic approaches [7, 2] or convex relaxation techniques [4, 10, 15, 3]. The underlying energies typically take into account local color information and aim at grouping regions of coherent color information, possibly enhanced with interactive user input indicating the rough location of objects of interest. By now, respective methods allow to separate objects of interest in rather challenging images, despite similar colors of object and background and strong variation of the illumination – see Figure 1.

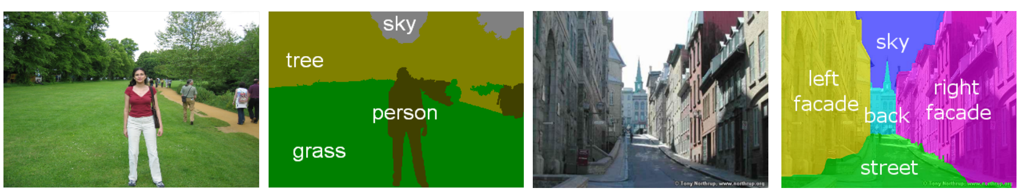


Input | Segmentation [13] | Input | Segmentation [8]

Figure 1: Interactive segmentations obtained using space-varying color models (left) and low-order moment constraints (right).

2. ...with Semantic Knowledge

Somewhat independent from the above developments in low-level image analysis, researchers have developed algorithms for high-level image analysis which allow to detect and recognize objects in images and even allow to perform an entire semantic scene analysis. Rather than modeling the color variations on a pixel-level they compute histograms of sparse features which are then related to respective features of previously observed objects [5]. Respective methods exhibit excellent performance on challenging high-level tasks. Yet the choice of features is generally somewhat heuristic and computed solutions typically do not come with a notion of statistical optimality with respect to the original image data, nor do they provide a per-pixel semantic decomposition of images.



Input | Segmentation [9] | Input | Segmentation [14]

Figure 2: Semantic segmentations obtained using label co-occurrence statistics (left) and ordering constraints (right).

In contrast, the framework of multilabel optimization allows to perform semantic image parsing on a per-pixel level with higher-level knowledge. Figure 2 shows recent examples where the multilabel optimization process was enhanced with a statistical prior on label co-occurrence [9] and with label ordering constraints [11, 6, 14]. In my view the fusion of low-level and high-level aspects of visual processing on the basis of efficient multi-label optimization methods bears great potential for future research in computer vision.

References

[1] Blake, A., Zisserman, A., 1987. Visual Reconstruction. MIT Press.

[2] Boykov, Y., Veksler, O., Zabih, R., 1998. Markov random fields with efficient approximations. In: Proc. IEEE Conf. on Comp. Vision Patt. Recog. (CVPR'98). Santa Barbara, California, pp. 648–655.

[3] Chambolle, A., Cremers, D., Pock, T., November 2008. A convex approach for computing minimal partitions. Technical report TR-2008-05, Dept. of Computer Science, University of Bonn, Bonn, Germany.

[4] Chan, T., Esedoḡlu, S., Nikolova, M., 2006. Algorithms for finding global minimizers of image segmentation and denoising models. SIAM Journal on Applied Mathematics 66 (5), 1632–1648.

[5] Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: Int. Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 886–893.

[6] Felzenszwalb, P. F., Veksler, O., 2010. Tiered scene labeling with dynamic programming. In: Int. Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 3097–3104.

[7] Greig, D. M., Porteous, B. T., Seheult, A. H., 1989. Exact maximum *a posteriori* estimation for binary images. J. Roy. Statist. Soc., Ser. B. 51 (2), 271–279.

[8] Klodt, M., Cremers, D., 2011. A convex framework for image segmentation with moment constraints. In: IEEE Int. Conf. on Computer Vision (ICCV).

[9] Ladicky, L., Russell, C., Kohli, P., Torr, P. H., 2008. Graph cut based inference with co-occurrence statistics. In: Europ. Conf. on Computer Vision.

[10] Lellman, J., Kappes, J., Yuan, J., Becker, F., Schnörr, C., October 2008. Convex multi-class image labeling by simplex-constrained total variation. Tech. rep., IPA, HCI, Dept. Of Mathematics and Computer Science, Univ. Heidelberg.

[11] Liu, X., Veksler, O., Samarabandu, J., 2010. Order-preserving moves for graph-cut-based optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 1182–1196.

[12] Mumford, D., Shah, J., 1989. Optimal approximations by piecewise smooth functions and associated variational problems. Comm. Pure Appl. Math. 42, 577–685.

[13] Nieuwenhuis, C., Toeppe, E., Cremers, D., 2011. Space-varying color distributions for interactive multiregion segmentation: Discrete versus continuous approaches. In: Int. Conf. on Energy Minimization Methods for Computer Vision and Pattern Recognition.

[14] Strekalovskiy, E., Cremers, D., 2011. Generalized ordering constraints for multilabel optimization. In: IEEE Int. Conf. on Computer Vision (ICCV).

[15] Zach, C., Gallup, D., Frahm, J.-M., Niethammer, M., October 2008. Fast global labeling for real-time stereo using multiple plane sweeps. In: Workshop on Vision, Modeling and Visualization.