Vertebral Body Segmentation in MRI via Convex Relaxation and Distribution Matching

Ismail Ben Ayed[1], Kumaradevan Punithakumar[1], Rashid Minhas[1], Rohit Joshi[2], and Gregory J. Garvin[2]

[1] GE Healthcare, London, ON, Canada
[2] The University of Western Ontario, London, ON, Canada

Abstract. We state vertebral body (VB) segmentation in MRI as a distribution-matching problem, and propose a convex-relaxation solution which is amenable to parallel computations. The proposed algorithm does not require a complex learning from a large manually-built training set, as is the case of the existing methods. From a very simple user input, which amounts to only three points for a whole volume, we compute a multi-dimensional model distribution of features that encode contextual information about the VBs. Then, we optimize a functional containing (1) a feature-based constraint which evaluates a similarity between distributions, and (2) a total-variation constraint which favors smooth surfaces. Our formulation leads to a challenging problem which is not directly amenable to convex-optimization techniques. To obtain a solution efficiently, we split the problem into a sequence of sub-problems, each can be solved exactly and globally via a convex relaxation and the augmented Lagrangian method. Our parallelized implementation on a graphics processing unit (GPU) demonstrates that the proposed solution can bring a substantial speed-up of more than 30 times for a typical 3D spine MRI volume. We report quantitative performance evaluations over 15 subjects, and demonstrate that the results correlate well with independent manual segmentations.

1 Introduction

Precise segmentation of the vertebral bodies (VBs) and intervertebral discs (IVDs) in MRI is an essential step towards thorough, reproducible and fast diagnosis of spine deformities [7,11]. Unlike CT, MRI scans depict soft-tissue structures, thereby allowing to characterize/quantify common spine disorders such as herniation [6] and disc degeneration [11]. However, most of related spine-segmentation works focused on CT, mainly because the latter affords high contrast for bony structures, e.g., [8,10,16], among others. In MRI, the problem is more challenging because of the intensity similarities and weak edges between the VBs and IVDs, the strong noise which results in intensity inhomogeneity within the VBs, and numerous acquisition protocols with different resolutions and noise types. Based on standard techniques such as adaptive-boosting learning combined with normalized cuts [7], active shape models [15], and fuzzy clustering

N. Ayache et al. (Eds.): MICCAI 2012, Part I, LNCS 7510, pp. 520–527, 2012.

combined with atlas registration [11], existing MRI-based spine segmentation algorithms require an intensive learning from a large, manually-segmented training set and costly pose-estimation procedures. Furthermore, in some cases, they are difficult to extend beyond the 2D case, e.g., [7,11]. Although effective in some cases, training-based algorithms may have difficulty in capturing the substantial variations in a clinical context. The ensuing results are often bounded to the choice of a training set and a specific type of MRI data.

In this study, we state VB segmentation in MRI as a distribution-matching problem, and propose a convex-relaxation solution which is amenable to parallel computations. From a very simple user input, which amounts to only three points (clicks) for a whole 3D volume (cf. the example in Fig. 1), we compute a multi-dimensional model distribution of features that encode contextual information about the VBs and their neighboring structures. Then, we optimize a functional containing (1) a feature-based constraint which evaluates a similarity between distributions, and (2) a total-variation constraint which favors smooth surfaces. Our formulation leads to a challenging problem which is not directly amenable to convex-optimization techniques. We split the problem into a sequence of sub-problems, each can be solved globally via a convex relaxation and the augmented Lagrangian method. Unlike related graph-cut approaches [1,13], the proposed convex-relaxation solution can be parallelized to reduce substantially the computational time for 3D domains (or higher), extends directly to high dimensions, and does not have the grid-bias problem. The proposed algorithm does not require a complex learning from a large manually-segmented training set, as is the case of the existing spine-segmentation methods. Therefore, the ensuing results are independent of the choice of a training set and a specific type of MRI data. We report quantitative performance evaluations over 15 subjects, and demonstrate that the results correlate well with independent manual segmentations. Our parallelized implementation on a graphics processing unit (GPU) demonstrates that the proposed solution can bring a substantial speed-up of more than 30 times for a typical 3D spine MRI volume.

Distribution-matching formulations have recently attracted a significant interest in computer vision [1,12,13]. Several studies have shown that, in the context of 2D color segmentation [1,13], such global constraints can yield outstanding performances unattainable with standard segmentation algorithms. However, these works are based on either active contours [2,12] or iterative graph cuts [1,13]. The active contour solutions were obtained following standard gradient-descent procedures [12], which lead to computationally intensive algorithms [1,12], more so when the image dimension is high (3D or higher). The contour evolution ensuing from a distribution measure is incremental, and requires a large number of updates of computationally expensive integrals [1,12]. The recent graph cut solutions in [1,13] demonstrated significant improvements over active contours in regard to computational load/speed as well as optimality of the solution. Unfortunately, graph cuts are not amenable to parallel computations [18]. In practice, it is well known that graph cuts can yield an excellent performance in the case of 2D grids with 4-neighborhood systems [4]. However, the efficiency may decrease

considerably when moving from 2D to 3D (or higher-dimensional) grids and when using larger-neighborhood grids [9]. Furthermore, the well-known grid bias is another limitation of graph-based approaches [14].

2 Formulation

Let $\mathbf{F} : \Omega \subset \mathbb{R}^n \rightarrow \mathcal{Z} \subset \mathbb{R}^k$ be a function which maps 3D (or higher-dimensional) domain Ω to a multi-dimensional (k-dimensional) space of contextual features $\mathcal{Z}$. For each point $\mathbf{x} \in \Omega$, $\mathbf{F}(\mathbf{x})$ is a vector containing image statistics within several box-shaped image patches of different orientations/scales (refer to the illustration in the left-hand side of Fig. 2). Such patch-based features can encode contextual knowledge about the region of interest and its neighboring structures (e.g., size, shape, orientation, relationships to neighboring structures, etc.). Let $\mathcal{M}$ denotes a k-dimensional model distribution of features learned from $\mathbf{F}$ within a rectangular approximation of one VB in a single 2D mid-sagittal slice of Ω (refer to the example in Fig. 1). Such approximation is obtained from a very simple user input which amounts to only three points (clicks). Given $\mathcal{M}$, we state VB segmentation as a distribution-matching problem. Our objective is to find an optimal region in Ω, so that (1) the distribution of the contextual features within the region most closely matches model $\mathcal{M}$ and (2) the surface of the region is smooth. We solve the following optimization problem:

$$\min_{u \in \{0,1\}} E(u) = \underbrace{-\sum_{\mathbf{z} \in \mathcal{Z}} \sqrt{\mathbf{P}_u(\mathbf{z}) \mathcal{M}(\mathbf{z})}}_{Contextual\ \ Distribution\ \ \ matching} + \lambda \underbrace{\int_{\Omega} |\nabla u(\mathbf{x})|\, d\mathbf{x}}_{Smoothness} \quad \text{where}$$

$$\mathbf{P}_u(\mathbf{z}) = \frac{\int_{\Omega} u(\mathbf{x}) K(\mathbf{F}(\mathbf{x}) - \mathbf{z}) d\mathbf{x}}{\int_{\Omega} u(\mathbf{x}) d\mathbf{x}} \text{ and } K(\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} exp^{-\frac{\|\mathbf{y}\|^2}{2\sigma^2}} \quad (1)$$

$u : \Omega \rightarrow \{0, 1\}$ is binary function which defines a variable partition of Ω: $\{\mathbf{x} \in \Omega / u(\mathbf{x}) = 1\}$, corresponding to the target region, and $\{\mathbf{x} \in \Omega / u(\mathbf{x}) = 0\}$, corresponding to the complement of the target region in Ω. $\mathbf{P}_u$ is the kernel density estimate (KDE) of the k-dimensional distribution of features $\mathbf{F}$ within variable region $\{\mathbf{x} \in \Omega / u(x) = 1\}$. K is a Gaussian kernel (σ is the width of the kernel). The distribution-matching term in (1) measures the Bhattacharyya similarity between $\mathbf{P}_u$ and $\mathcal{M}$. This measure has a fixed (normalized) range which affords a conveniently practical appraisal of the similarity. This is not the case for the other common measures (e.g., the Kullback-Leibler divergence).

The smoothness term is a standard total-variation constraint which penalizes the occurrences of small, isolated regions in the solution. λ is a positive constant that balances the contribution of each term.

Direct computation of (1) is a challenging problem due to the high non-convexity of the distribution-matching term. To obtain a solution efficiently, we split the problem into a sequence of sub-problems, each of which can be solved globally via a convex relaxation and the augmented Lagrangian method.

Sub-problems: Rather than optimizing directly E, we optimize iteratively a sequence of instrumental functions, denoted $A(u, u^i)$, $i \geq 1$ (i is the iteration number), whose optimization is easier than E:

$$u^{i+1} = \min_{u \in \{0,1\}} A(u, u^i), \quad i \geq 1 \quad \text{s.t.} \tag{2a}$$

$$E(u) \leq A(u, u^i), \quad i \geq 1 \tag{2b}$$

$$E(u) = A(u, u) \quad \forall u : \Omega \to \{0, 1\} \tag{2c}$$

Using the constraints in (2b) and (2c), and by definition of minimum in (2a), one can show that the sequence of solutions in (2a) yields a decreasing sequence of E: $E(u^i) = A(u^i, u^i) \geq A(u^{i+1}, u^i) \geq E(u^{i+1})$. Furthermore, $E(u^i)$ is lower bounded and, therefore, converges to a minimum of E. Now, consider the following proposition [1]:

Proposition 1. *Given a fixed u^i, for any $u : \Omega \to \{0, 1\}$ verifying $\{\mathbf{x} \in \Omega / u(\mathbf{x}) = 1\} \subset \{\mathbf{x} \in \Omega / u^i(\mathbf{x}) = 1\}$ and $\forall \alpha \in [0, \frac{1}{2}]$, the following function verifies the constraints in (2b) and (2c) and, therefore, its iterative optimization yields a minimum of E (in the expression of A_α, $\mathbf{x}$ and $\mathbf{z}$ are omitted as arguments to simplify the notations):*

$$A_\alpha(u, u^i) = -(1+\alpha) \sum_{\mathbf{z} \in \mathcal{Z}} \sqrt{\mathbf{P}_{u^i} \mathcal{M}} + \int_\Omega \{\alpha u g^i + (1+\alpha)(1-u) h^i + \lambda |\nabla u|\} d\mathbf{x}$$

$$\text{where } g^i = \frac{\sum_{\mathbf{z} \in \mathcal{Z}} \sqrt{\mathbf{P}_{u^i}(\mathbf{z}) \mathcal{M}(\mathbf{z})}}{\int_\Omega u^i d\mathbf{x}}; \quad h^i = \frac{u^i}{\int_\Omega u^i d\mathbf{x}} \sum_{\mathbf{z} \in \mathcal{Z}} K(\mathbf{z} - \mathbf{F}) \sqrt{\frac{\mathcal{M}(\mathbf{z})}{\mathbf{P}_{u^i}(\mathbf{z})}} \tag{3}$$

Convex Relaxation and Equivalent Constrained Problem: Now, we minimize A_α over $u \in \{0, 1\}$ via convex optimization. However, $\min_{u \in \{0,1\}} A_\alpha$ is still non-convex due to the binary-valued constraint $u \in \{0, 1\}$. We first relax such constraint to interval $[0, 1]$, thereby obtaining the following convex problem: $\min_{u \in [0,1]} A_\alpha$. Then, we propose the following result which allows us to obtain an exact and global minimum of A_α over $u \in \{0, 1\}$ via the multiplier-based augmented Lagrangian method (the proof follows the ideas of [17,18]):

Proposition 2. *The convex problem $\min_{u \in [0,1]} A_\alpha(u, u^i)$ is equivalent to the following constrained problem:*

$$\max_{p_h, p_g, p} \min_{u} \int_\Omega \{p_h + u(\operatorname{div} p - p_h + p_g)\} d\mathbf{x} \quad s.t.$$

$$p_h(\mathbf{x}) \leq (1+\alpha) h^i(\mathbf{x}); \quad p_g(\mathbf{x}) \leq \alpha g^i(\mathbf{x}); \text{ and } |p(\mathbf{x})| \leq \lambda \text{ a.e. } \mathbf{x} \in \Omega \tag{4}$$

where u is viewed as the multiplier to constraint $\operatorname{div} p - p_h + p_g = 0$. $p : \Omega \to \mathbb{R}$, $p_h : \Omega \to \mathbb{R}$ and $p_g : \Omega \to \mathbb{R}$ are variables in the form of scalar functions. Furthermore, by simply thresholding the optimum $u^* \in [0, 1]$ of (4), we obtain an exact and global optimum of the non-convex problem $min_{u \in \{0,1\}} A_\alpha(u, u^i)$.

[1] The proof is given in the supplemental material available at this link: `http://externe.emt.inrs.ca/users/benayedi/BenAyed-Miccai12-Supp.pdf`

The equivalence in proposition 2 is similar to the equivalence of the continuous max-flow/min-cut problem in [17,18]. This equivalence allows us to derive an efficient multiplier-based algorithm for optimizing E. The algorithm is based on the standard augmented Lagrangian method [3,17,18]. We define the following augmented Lagrangian function corresponding to the problem in (4) (multiplier u is replaced here by v to avoid confusion between inner and outer iterations in the algorithm presented next): $L_c(p_h, p_g, p, v) = \int_\Omega \{p_h + v(\operatorname{div} p - p_h + p_g) - \frac{c}{2} \|\operatorname{div} p - p_h + p_g\|^2\} d\mathbf{x}$. c is a positive constant. Here following a summary of the algorithm (i and j are the numbers of inner end outer iterations respectively).

Algorithm 1: Multiplier-based augmented-Lagrangian optimization

- Initialize u by $u^0(x) = 1$ $(i = 0)$
- **repeat**
  1. Update g^i and h^i according to (3)
  2. set $j = 0$, $p = p^0$, $p_h = p_h^0$, $p_g = p_g^0$ and $v = v^0$
  3. **repeat**
     - Optimize L_c with respect p:
       $p^{j+1} = \max_{|p| \leq \lambda} -\frac{c}{2} \left\| \operatorname{div} p - p_h^j + p_g^j - \frac{v^j}{c} \right\|^2$
       This can be solved by the Chambolle's algorithm [5].
     - Optimize L_c with respect to p_h (closed-form solution):
       $p_h^{j+1} = \max_{p_h < (1+\alpha) h^i} \int_\Omega p_h - \frac{c}{2} \left\| \operatorname{div} p^{j+1} - p_h + p_g^j - \frac{v^j}{c} \right\|^2 dx$
     - Optimize L_c with respect p_g (closed-form solution):
       $p_g^{j+1} = \max_{p_g \leq \alpha g^i} -\frac{c}{2} \left\| \operatorname{div} p^{j+1} - p_h^{j+1} + p_g - \frac{v^j}{c} \right\|^2$
     - Update multiplier v:
       $v^{j+1} = v^j - c(\operatorname{div} p^{j+1} - p_h^{j+1} + p_g^{j+1})$
     - Let $j = j + 1$

  until *Convergence*;
  4. Let $v*$ the solution obtained from the inner iterations above. Compute a binary solution by applying Otsu's thresholding to $v*$.
  5. Let $i = i + 1$
  6. Let u^i equal to the binary solution obtained at step 4.

  until *Convergence*;

3 Experiments

Description of the Features: The left-hand side of Fig. 2 depicts the contextual features we used in our experiments. For each point $\mathbf{x} \in \Omega$, we built a feature vector of dimension 3: $\mathbf{F}(\mathbf{x}) = (\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3)$, with $\mathbf{F}_1$ the mean of intensity within a 21 **x** 7 **x** 1 rectangular-shaped, vertically-oriented patch, $\mathbf{F}_2$ the mean of intensity within a 7 **x** 21 **x** 1 rectangular-shaped, horizontally-oriented patch, and $\mathbf{F}_3$ the mean intensity within a 7 **x** 7 **x** 1 square-shaped patch, all centered at point **x**. We used 32 bins to compute the feature distributions. Note that other features based on image gradients or texture can be used within the same framework.

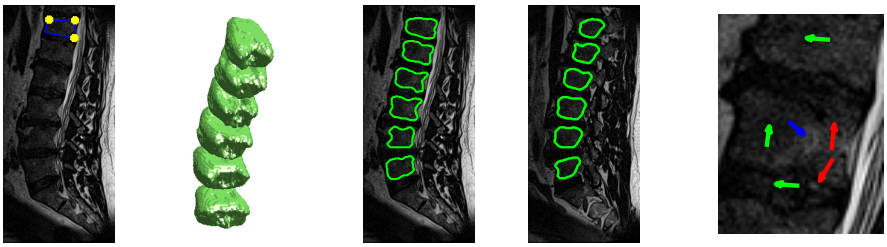

Fig. 1. A typical 3D example with a 320 **x** 150 **x** 80 volume. First column: three-point user input in a single 2D mid-sagittal slice (slice 35); second column: the obtained 3D surfaces of the lumbar (L1 to L5) and T12 VBs; third and fourth columns: the corresponding 2D-slice results (slices 35 and 25). Last column: an image patch from the user-input slice, which illustrates the difficulties inherent to spine MRI segmentation. The smoothness weight is set equal to one ($\lambda = 1$).

A Typical 3D Example: Fig. 1 depicts a typical example of the results using a 320 **x** 150 **x** 80 lumbar spine volume of the type T2-weighted. The first column shows the simple, three-point user input in a single 2D mid-sagittal slice (slice 35). The second column shows the obtained result, depicted by the 3D surfaces of the lumbar (L1 to L5) and T12 VBs. The third and fourth columns depict the corresponding 2D-slice results (slices 35 and 25). The last column is an image patch from the user-input slice, which illustrates the difficulties inherent to spine MRI segmentation. The red arrows point to weak edges between the VBs and IVDs, the green arrows to intensity similarities between the VBs and IVDs, and the blue arrow to strong image noise within the VBs.

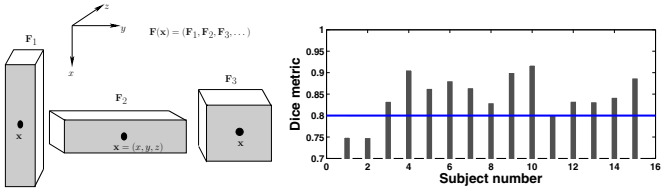

Fig. 2. Left: illustration of the contextual features; right: *DM* evaluations

Computational Evaluations: We further evaluated two implementations (with and without parallelization), one GPU based and the other CPU (central processing unit) based. The parallelized computations were run on an NVIDIA Quadro FX3700 with 112 Cuda cores, whereas the non-parallelized version was run on an E5440 quad core 2.83 GHz Xeon, with 3.25GB of RAM. Both versions were implemented in C. Table 1 reports the GPU/CPU times corresponding to the example in Fig. 1. For the 3D case, the parallelized version brought a

substantial speed-up of more than 30 times. In the 2D case, the GPU version brought a less significant, but nonetheless important, speed-up of 13 times.

Table 1. Computational GPU/CPU times of the proposed solution

Implementation ($\lambda = 1$)	3D (320 **x** 150 **x** 80)	2D (320 **x** 150)
GPU (Parallelized)	143.08 s ($<$ 3 min)	0.64 s
CPU	$4.50\mathbf{x}10^3$ ($\approx$ 75 min)	8.46 s

Table 2. Quantitative performance evaluations over 15 subjects

DM mean	*DM* std	*Correlation coefficient (r)*
0.85	0.051	0.98

Quantitative Evaluations: The evaluation was carried out over a data set of 15 mid-sagittal 2D MR spine scans acquired from 15 different subjects. We segmented automatically a total of 75 lumbar VBs starting from a simple three-point user input. We focused on lumbar VBs and MRI data of the type T2-weighted. However, the proposed method can be readily extended to IVDs and other MRI types, as it does not require any shape or image specific training. The results were compared to independent manual segmentations approved by an expert. We assessed the similarities between the ground truth and the obtained segmentations using two measures: the Dice metric (DM) and the correlation coefficient (r). DM is commonly used to measure the similarity (overlap) between the automatically detected and ground-truth regions: $DM = \frac{2\mathbf{S_{am}}}{\mathbf{S_a}+\mathbf{S_m}}$, with $\mathbf{S_a}$, $\mathbf{S_m}$, and $\mathbf{S_{am}}$ corresponding respectively to the sizes of the segmented lumbar region (i.e., the region containing all the VBs from L1 to L5), the corresponding hand-labeled region, and the intersection between them. Table 2 reports for all the data analyzed the DM mean and standard deviation, as well as the correlation coefficient between manual and automatic region sizes. Fig. 2 (right) depicts the DM for the 15 analyzed lumbar regions (each region contains 5 VBs). We obtained a $DM > 0.80$ for 13 subjects (DM is about 0.75 for only two subjects). The proposed method also yielded a high correlation: $r = 0.98$.

4 Conclusion

We proposed a distribution-matching algorithm for VB segmentation in MRI and a convex-relaxation solution which is amenable to parallel computations. The algorithm removes the need for a complex learning from a large training set. We described quantitative performance evaluations over 15 subjects, and showed that a GPU-based implementation of our solution can bring a substantial speed-up of more than 30 times for a typical 3D spine MRI volume.

References

1. Ben Ayed, I., Chen, H.M., Punithakumar, K., Ross, I., Li, S.: Graph cut segmentation with a global constraint: Recovering region distribution via a bound of the bhattacharyya measure. In: CVPR, pp. 3288–3295 (2010)
2. Ben Ayed, I., Li, S., Ross, I.: A statistical overlap prior for variational image segmentation. International Journal of Computer Vision 85(1), 115–132 (2009)
3. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific (1999)
4. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. IEEE Trans. on Pattern Analysis and Machine Intelligence 26(9), 1124–1137 (2004)
5. Chambolle, A.: An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision 20(1-2), 89–97 (2004)
6. Fardon, D.F., Milette, P.C.: Nomenclature and classification of lumbar disc pathology: Recommendations of the combined task forces of the north american spine society, american society of spine radiology, and american society of neuroradiology. spine 26(5), E93–E113 (2001)
7. Huang, S., Chu, Y., Lai, S., Novak, C.: Learning-based vertebra detection and iterative normalized-cut segmentation for spinal mri. IEEE Trans. on Medical Imaging 28(10), 1595–1605 (2009)
8. Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., Lorenz, C.: Automated model-based vertebra detection, identification, and segmentation in ct images. Medical Image Analysis 13(3), 471–482 (2009)
9. Klodt, M., Schoenemann, T., Kolev, K., Schikora, M., Cremers, D.: An Experimental Comparison of Discrete and Continuous Shape Optimization Methods. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 332–345. Springer, Heidelberg (2008)
10. Mastmeyer, A., Engelke, K., Fuchs, C., Kalender, W.: A hierarchical 3d segmentation method and the definition of vertebral body coordinate systems for qct of the lumbar spine. Medical Image Analysis 10(4), 560–577 (2006)
11. Michopoulou, S., Costaridou, L., Panagiotopoulos, E., Speller, R., Panayiotakis, G., Todd-Pokropek, A.: Atlas-based segmentation of degenerated lumbar intervertebral discs from mr images of the spine. IEEE Trans. on Biomedical Engineering 56(9), 2225–2231 (2009)
12. Mitiche, A., Ben Ayed, I.: Variational and Level Set Methods in Image Segmentation, 1st edn. Springer (2010)
13. Pham, V.Q., Takahashi, K., Naemura, T.: Foreground-background segmentation using iterated distribution matching. In: CVPR (2011)
14. Pock, T., Chambolle, A., Cremers, D., Bischof, H.: A convex relaxation approach for computing minimal partitions. In: CVPR (2009)
15. Seifert, S., Wachter, I., Schmelzle, G., Dillmann, R.: A knowledge-based approach to soft tissue reconstruction of the cervical spine. IEEE Transactions on Medical Imaging 28(4), 494–507 (2009)
16. Shen, H., Litvin, A., Alvino, C.: Localized Priors for the Precise Segmentation of Individual Vertebras from CT Volume Data. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 367–375. Springer, Heidelberg (2008)
17. Yuan, J., Bae, E., Tai, X.C., Boykov, Y.: A study on continuous max-flow and min-cut approaches. Part I: Binary labeling. Tech report CAM-10-61, UCLA (2010)
18. Yuan, J., Bae, E., Tai, X.C.: A study on continuous max-flow and min-cut approaches. In: CVPR (2010)