

Transfer Shape Modeling Towards High-Throughput Microscopy Image Segmentation

Fuyong Xing^{1,2}, Xiaoshuang Shi², Zizhao Zhang³, JinZheng Cai², Yuanpu Xie², and Lin Yang^{1,2,3}(✉)

¹ Department of Electrical and Computer Engineering,
University of Florida, Gainesville, FL 32611, USA

² J. Crayton Pruitt Family Department of Biomedical Engineering,
University of Florida, Gainesville, FL 32611, USA

lin.yang@bme.ufl.edu

³ Department of Computer and Information Science and Engineering,
University of Florida, Gainesville, FL 32611, USA

Abstract. In order to deal with ambiguous image appearances in cell segmentation, high-level shape modeling has been introduced to delineate cell boundaries. However, shape modeling usually requires sufficient annotated training shapes, which are often labor intensive or unavailable. Meanwhile, when applying the model to different datasets, it is necessary to repeat the tedious annotation process to generate enough training data, and this will significantly limit the applicability of the model. In this paper, we propose to transfer shape modeling learned from an existing but different dataset (e.g. lung cancer) to assist cell segmentation in a new target dataset (e.g. skeletal muscle) without expensive manual annotations. Considering the intrinsic geometry structure of cell shapes, we incorporate the shape transfer model into a sparse representation framework with a manifold embedding constraint, and provide an efficient algorithm to solve the optimization problem. The proposed algorithm is tested on multiple microscopy image datasets with different tissue and staining preparations, and the experiments demonstrate its effectiveness.

1 Introduction

Automatic cell segmentation is a critical step in microscopy image analysis, and serves as a basis of many subsequent quantitative analyses [4], such as cellular morphology calculation and individual cell classification. It is challenging to achieve robust cell segmentation due to ambiguous image appearance, such as weak cell boundaries, inhomogeneous intracellular intensity, and partial occlusion of cells, etc. Therefore, instead of purely relying on low-level image appearances, high-level shape modeling has been introduced to improve object boundary delineation [2, 15, 18, 19].

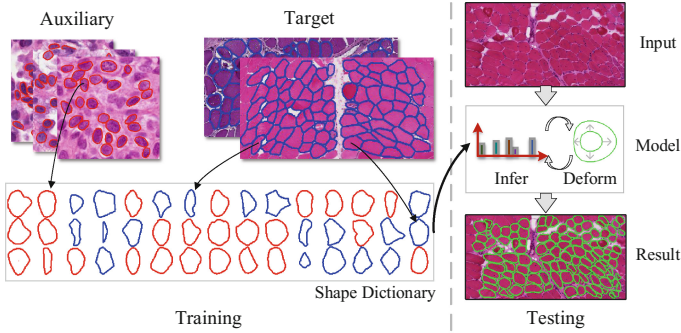


Fig. 1. The overview of the proposed shape transfer model. It aims at learning a cell shape dictionary, which can produce data representations applicable to the target set.

Effective shape modeling usually requires a sufficient number of annotated training data, which might be labor intensive or even unavailable in some applications. Meanwhile, it is necessary to repeat the tedious annotation process to generate training shapes for different microscopy image datasets, thereby significantly limiting the generality of shape modeling. In this paper, we propose to transfer shape modeling from an auxiliary dataset with sufficient training cell shapes to a target dataset with a limited number of training samples for automatic cell segmentation. In order to respect the intrinsic geometry structure of cell shapes [16], we incorporate the shape transfer model into a sparse representation framework with a manifold embedding constraint. Cell shape transfer is modeled as learning a compact dictionary that can construct robust representations for cells from different datasets, and model optimization is achieved by using an efficient sparse encoding algorithm. In this scenario, we can significantly reduce the manual annotation efforts and improve the generality of the shape modeling across multiple datasets so as to produce high-throughput cell segmentation in microscopy image analysis.

2 Methodology

An overview of the proposed shape transfer modeling for cell segmentation is shown in Fig. 1. In the training stage, the shape dictionary is learned by regularizing the distribution differences of cell training shapes between the auxiliary and target datasets, and thus the dictionary can serve as a reference repository to generate robust data representations across both datasets. Convolutional neural networks (CNNs) [5, 15] are learned to conduct pixel-wise classification for generating initial cell contours/shapes. In the testing stage, the framework moves the contours towards cell boundaries until convergence in an alternate manner [18]: deform shapes with an efficient active contour model [20] and infer shapes with the transferred shape priors based on the learned dictionary. Due to page limits, this paper only focuses on shape transfer modeling.

2.1 Shape Transfer Modeling

Problem Definition: Denote auxiliary data by $\mathcal{D}_a = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_a}\}$ with N_a training cell shapes and target data by $\mathcal{D}_t = \{\mathbf{x}_{N_a+1}, \mathbf{x}_{N_a+2}, \dots, \mathbf{x}_{N_a+N_t}\}$ with N_t training shapes, where each shape is described by the concatenated 2D coordinates of p evenly-sampled control points and aligned by removing global transformation. Based on \mathcal{D}_a and \mathcal{D}_t , our goal is to learn a compact shape dictionary $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_K] \in \mathbb{R}^{2p \times K}$ such that for any cell shape $\mathbf{x} \in \mathbb{R}^{2p \times 1}$, a solution to $\mathbf{x} = \mathbf{B}\boldsymbol{\alpha}$, with sparse constraints on the coefficient $\boldsymbol{\alpha}$, is robust and effective across both datasets. In this way, the cell shape information from the auxiliary dataset would be transferred and thus assist the data representation generation for cell segmentation in the target dataset.

Shape Transfer Model: Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{2p \times N}$ with $N = N_a + N_t$ be the input data matrix consisting of \mathcal{D}_a and \mathcal{D}_t , sparse shape modeling aims to learn a dictionary by minimizing the reconstruction error with specific constraints on the coefficients [18, 19]. In this paper, we model dictionary learning as a subset selection problem which seeks a small set of representatives to summarize and describe the whole dataset \mathbf{X} , thereby removing outliers that are not true representatives and improving the runtime computational efficiency. A straightforward way to conduct sparse subset selection is to regularize the coefficient matrix with an $\ell_{2,1}$ or $\ell_{\infty,1}$ norm [6]. Meanwhile, since the number of constraints for shape control is limited, cell shapes actually lie on a low-dimensional manifold [16]. Therefore, we can formulate the subset selection as a graph and row-sparsity regularized optimization problem

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{X}\mathbf{A}\|_F^2 + \gamma \text{Tr}(\mathbf{A}\mathbf{L}\mathbf{A}^T) + \lambda \|\mathbf{A}\|_{2,1}, \text{ s.t. } \mathbf{1}_N^T \mathbf{A} = \mathbf{1}_N, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the sparse coefficient matrix. $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the graph Laplacian, where \mathbf{W} is a n -nearest neighbor graph ($n = 5$) with nodes representing the training shapes: $W_{ij} = 1$ if \mathbf{x}_i is among the n -nearest neighbors of \mathbf{x}_j , otherwise 0; \mathbf{D} is a diagonal matrix with $D_{ii} = \sum_{j=1}^N W_{ij}$. $\text{Tr}(\cdot)$ denotes the trace operation, and it encourages those adjacent shapes in the intrinsic geometry to be represented with similar codes. $\|\mathbf{A}\|_{2,1} = \sum_{i=1}^N \|\boldsymbol{\alpha}^i\|_2$ ($\boldsymbol{\alpha}^i$ representing the i -th row of \mathbf{A}) is the sum of the ℓ_2 norms of the rows and is a convex relaxation for counting the number of nonzero rows of \mathbf{A} . The affine constraint ensures shift invariance, and $\mathbf{1}_N \in \mathbb{R}^{N \times 1}$ is a vector with all elements equal to one. Since each row $\boldsymbol{\alpha}^i$ corresponds to one input cell shape \mathbf{x}_i , after solving (1) we can simply form the dictionary $\mathbf{B} = [\mathbf{x}_{b_1}, \dots, \mathbf{x}_{b_K}]$ by selecting the representatives corresponding to the nonzero rows of \mathbf{A} and apply it to cell shape encoding based on sparse reconstruction in the testing stage.

In order to transfer the shape prior knowledge from the auxiliary to target dataset, the sparse encoding needs to be robust across both datasets. An intuitive strategy is to enable the selected representatives or the dictionary to capture the common composition of the two datasets instead of only the individual characteristics of the auxiliary shapes, and this can be realized by reducing the distribution differences of the representations between two datasets. Inspired

by [8, 11], we propose to penalize the distance in Maximum Mean Discrepancy (MMD), which is a nonparametric criterion to measure the distribution difference between the means of samples from two datasets in a transformed representation space. In our model, cell shapes are mapped into a representation space via the learned dictionary, and thus we need to penalize the distance between the auxiliary and target datasets in the sparse coefficients

$$\left\| \frac{1}{N_a} \sum_{i=1}^{N_a} \boldsymbol{\alpha}_i - \frac{1}{N_t} \sum_{j=N_a+1}^{N_a+N_t} \boldsymbol{\alpha}_j \right\|^2 = \sum_{i,j=1}^{N_a+N_t} \boldsymbol{\alpha}_i^T \boldsymbol{\alpha}_j M_{ij} = \text{Tr}(\mathbf{A} \mathbf{M} \mathbf{A}^T), \quad (2)$$

where $\boldsymbol{\alpha}_i \in \mathbb{R}^{N \times 1}$ is the i -th column of \mathbf{A} , and $\mathbf{M} \in \mathbb{R}^{N \times N}$ is the MMD matrix with the ij -th element computed as $M_{ij} = \frac{1}{N_a^2}$ when $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_a$; $M_{ij} = \frac{1}{N_a N_t}$ when $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_t$ and $M_{ij} = \frac{-1}{N_a N_t}$ otherwise.

One of significant benefits for choosing the MMD criterion is that it does not require an intermediate density estimate, which is usually a non-trivial task. By incorporating (2) into (1), we obtain the proposed objective function for cell shape knowledge transfer

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{X} \mathbf{A}\|_F^2 + \gamma \text{Tr}(\mathbf{A} \mathbf{L} \mathbf{A}^T) + \mu \text{Tr}(\mathbf{A} \mathbf{M} \mathbf{A}^T) + \lambda \|\mathbf{A}\|_{2,1}, \text{ s.t. } \mathbf{1}_N^T \mathbf{A} = \mathbf{1}_N, \quad (3)$$

where $\mu > 0$ is a weight parameter controlling the MMD regularization. MMD asymptotically approaches to zero if the auxiliary and target datasets exhibit the same distribution [8]. By mapping cell shapes into a common representation space, solving (3) can refine the subset selection of representatives and transfer the shape knowledge of the auxiliary dataset, thereby producing representations that are applicable to the target dataset.

2.2 Efficient Local Encoding

The model in (3) can be solved by the Alternating Direction Method of Multipliers (ADMM) framework [3]. However, ADMM might introduce additional parameters in the optimization procedure. Note that the graph regularization in (3) respects the intrinsic Riemannian structure of cell shapes by obeying the manifold assumption: if \mathbf{x}_i and \mathbf{x}_j are adjacent in the manifold, then their representations, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\alpha}_j$, with respect to the learned dictionary should be close to each other. Actually this can be approximated in a more efficient way: using a globally linear function with respect to a set of learned local coordinates [17] and solving a much smaller linear system. Therefore, instead of directly penalizing the differences of the representations between neighboring shapes, during the sparse coding we can weight the codes with the similarity between the shapes and the dictionary bases. More importantly, the time complexity can be significantly reduced. Therefore, we can revise the model in (3) by replacing the graph regularization with a locality-constrained term and explicitly modeling the dictionary \mathbf{B}

$$\min_{\mathbf{B}, \tilde{\mathbf{A}}} J = \|\mathbf{X} - \mathbf{B} \tilde{\mathbf{A}}\|_F^2 + \gamma \sum_{i=1}^N \tilde{\boldsymbol{\alpha}}_i^T \mathbf{Q}^i \tilde{\boldsymbol{\alpha}}_i + \mu \text{Tr}(\tilde{\mathbf{A}} \mathbf{M} \tilde{\mathbf{A}}^T), \text{ s.t. } \mathbf{1}_K^T \tilde{\mathbf{A}} = \mathbf{1}_K, \quad (4)$$

where $\mathbf{Q}^i \in \mathbb{R}^{K \times K}$ is a diagonal matrix with $Q_{kk}^i = \|\mathbf{x}_i - \mathbf{b}_k\|^2$ representing the distance between shape \mathbf{x}_i and dictionary basis \mathbf{b}_k . The introduced locality constraint encourages each shape to be represented with its neighbors in \mathbf{B} . $\tilde{\mathbf{A}} \in \mathbb{R}^{K \times N}$ is the coefficient matrix. We remove the sparsity regularization in (4) since locality guarantees sparsity but not necessary vice versa [14, 17].

It is difficult to simultaneously compute the two unknown variables in (4), and thus we solve it in an alternate way: calculate coefficients $\tilde{\mathbf{A}}$ with dictionary \mathbf{B} fixed and learn dictionary \mathbf{B} with coefficients $\tilde{\mathbf{A}}$ fixed. We can derive the i -th coefficient analytically with a fixed dictionary as

$$\tilde{\alpha}_i = (\mathbf{Y} + \mu(M_{ii}\mathbf{I}_K + \mathbf{1}_K\mathbf{r}_i^T + \mathbf{r}_i\mathbf{1}_K^T) + \gamma\mathbf{Q}^i)^{-1}\mathbf{1}_K, \quad (5)$$

where $\mathbf{Y} = (\mathbf{1}_K\mathbf{x}_i^T - \mathbf{B}^T)(\mathbf{x}_i\mathbf{1}_K^T - \mathbf{B})$ and $\mathbf{r}_i = \sum_{j \neq i} M_{ij}\tilde{\alpha}_j$. In order to preserve the affine constraint, we further normalize $\tilde{\alpha}_i$ such that $\mathbf{1}_K^T\tilde{\alpha}_i = 1$. In our implementation, we anchor each shape in its local coordinate system for fast encoding.

With coefficients $\tilde{\mathbf{A}}$ fixed, dictionary \mathbf{B} is updated by gradient decent [10]. The derivative of the objective function with respect to the k -th basis \mathbf{b}_k is derived in (6). To ensure that the dictionary bases coincide with a subset of actual cell shapes, in each iteration we update the basis \mathbf{b}_k by selecting the shape \mathbf{x}_l that exhibits the largest correlation between the displacement and the negative gradient in (7)

$$\nabla J_{\mathbf{b}_k} = -2 \sum_{i=1}^N (\mathbf{x}_i - \mathbf{B}\tilde{\alpha}_i)\tilde{\alpha}_{ik} + (\mathbf{x}_i - \mathbf{b}_k)\tilde{\alpha}_{ik}^2, \quad (6)$$

$$\mathbf{b}_k = \arg \max_{\mathbf{x}_l \in \mathbf{X}} \frac{(\mathbf{x}_l - \mathbf{b}_k)^T (-\nabla J)}{\|\mathbf{x}_l - \mathbf{b}_k\|_2 \|-\nabla J\|_2}. \quad (7)$$

The dictionary basis update and the coefficient computation are alternately conducted until convergence. In the testing stage, a new cell shape \mathbf{x} from the target dataset can be encoded by solving (4) with a fixed \mathbf{B} and without the MMD term.

3 Experiments

Datasets and Experimental Setup: The proposed shape transfer model is extensively tested on multiple microscopy image datasets: lung cancer, pancreatic neuroendocrine tumor (NET), and skeletal muscle. The gold standards of cell boundaries are manually annotated. Lung cancer has over 20000 annotated cell shapes, and thus it is used as the auxiliary dataset with randomly selected about one-tenth for training. The target datasets include 45 NET images with half for training (the left for testing) and 41 skeletal muscle images with about three-fourths for training. The parameters in (4) are chosen as $\gamma = 0.001$ and $\mu = 10$. The dictionary size is chosen as one-tenth of the training data. Many metrics [9, 13] can be applied to quantitative analysis, and we choose three generic segmentation-based criteria [15]: Dice similarity coefficient (DSC), Hausdorff distance (HD), and mean absolute distance (MAD).

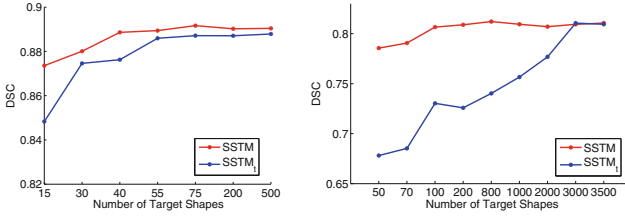


Fig. 2. Comparative segmentation accuracy (DSC) between the shape transfer model, SSTM, and the model trained with only target shapes, SSTM_t, on the NET (left) and muscle (right) datasets.

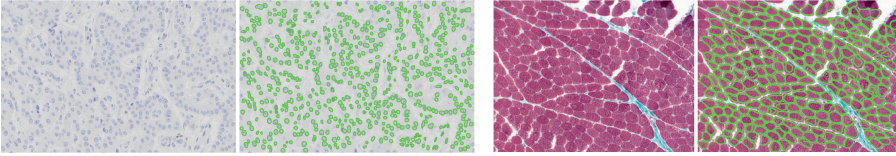


Fig. 3. Cell segmentation results on two sample images from the NET (left two) and muscle (right two) datasets. Cells touching image boundaries are ignored.

We train the CNN models by following [15] on the NET and [5] on the skeletal muscle images. For the former, the CNN model is trained with over 6×10^5 positive and 9×10^5 negative image patches. The parameters are with: total iterations of 2.5×10^5 , learning rate of 0.001, momentum of 0.8, and batch size of 128. For the latter, the model is trained with one million patches with half positives and half negatives. The parameters are with: total iterations of 2×10^5 , learning rate of 0.01, momentum of 0.9, weight decay of 0.1 (every 50000 iterations), and batch size of 256. CNN models provide coarse segmentation for subsequent contour deformation and refinement.

Shape Transfer Model Evaluation: Figure 2 shows the segmentation accuracy of the proposed model, SSTM, with respect to the number of target cell shapes and two hundred auxiliary data samples. The model learned using only target cell shapes (denoted by SSTM_t) is also provided for comparison. It is clear that when there exist limited target training samples, shape transfer modeling improves the segmentation accuracy and outperforms SSTM_t. When sufficient target shapes are applied to training, there exist no significant performance improvement using transfer shapes. Since muscle cells exhibit more significant shape variations than NET cells, SSTM_t requires much more target shapes to produce competitive performance to SSTM. We can see that when fixing a desired segmentation accuracy, transfer models require less target data. Figure 3 shows qualitative segmentation results on the two datasets, where many cells with weak boundaries are segmented.

Comparison with State of the Arts: We compare SSTM with four state of the arts: isoperimetric graph partition (IGP) [7], superpixel-based segmentation (SUP) [12], graph cut and coloring (GCC) [1], and repulsive level set (RLS) [13].

Table 1. Comparative pixel-wise segmentation accuracy on NET and muscle datasets. For each metric (DSC, HD, and MAD), the mean and standard deviation are listed.

	NET			Muscle		
	DSC	HD	MAD	DSC	HD	MAD
IGP [7]	0.49 ± 0.21	10.58 ± 14.07	8.16 ± 7.15	0.63 ± 0.24	12.81 ± 11.22	7.73 ± 5.62
SUP [12]	0.74 ± 0.18	7.30 ± 7.69	4.08 ± 3.76	0.63 ± 0.16	16.71 ± 9.52	8.86 ± 4.99
GCC [1]	0.60 ± 0.22	6.57 ± 4.38	5.18 ± 2.85	0.70 ± 0.22	12.47 ± 16.11	7.48 ± 9.13
RLS [13]	0.84 ± 0.09	4.25 ± 3.48	2.35 ± 2.11	0.84 ± 0.12	8.64 ± 10.53	4.97 ± 6.72
SSTM	0.89 ± 0.12	2.98 ± 3.14	2.01 ± 2.11	0.85 ± 0.13	4.88 ± 4.80	3.11 ± 2.63

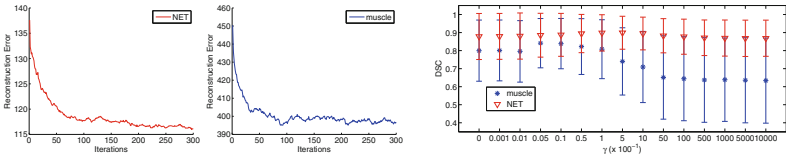


Fig. 4. Convergence study (left and middle) and parameter sensitivity analysis (right).

Table 1 lists the comparative performance on the two target datasets. As we can see, SSTM outperforms the others in terms of three segmentation criteria, especially in HD that measures the largest error for each segmentation. In addition, the lowest standard deviations in the metrics (for almost all cases) indicate the strong stability of the proposed approaches.

Convergence and Parameter Sensitivity Analysis: The reconstruction errors for model training with respect to the number of iterations are shown in Fig. 4, which indicates that the algorithm can converge in a limited number of iterations. We set $\mu = \{10, 100, 1000\}$ and find that there exist no statistically significant variations on the accuracy. The effects of parameter γ on the performance is provided in the right panel of Fig. 4, which shows our algorithm can achieve stable performance within a wide range of values.

4 Conclusion

In this paper, we propose a shape transfer model for cell segmentation in microscopy images. By learning a compact shape dictionary using an auxiliary dataset, it can generate shape representations that are applicable to the target dataset. In this scenario, it can significantly reduce expensive manual annotations of target training data. Extensive experiments on multiple microscopy image datasets demonstrate the effectiveness of the proposed method.

References

1. Al-Kofahi, Y., Lassoued, W., Lee, W., Roysam, B.: Improved automatic detection and segmentation of cell nuclei in histopathology images. *TBME* **57**(4), 841–852 (2010)

2. Ali, S., Madabhushi, A.: An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *TMI* **31**(7), 1448–1460 (2012)
3. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1–122 (2011)
4. Chang, H., Han, J., Borowsky, A., Loss, L., Gray, J., Spellman, P., Parvin, B.: Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association. *TMI* **32**(4), 670–682 (2013)
5. Cireşan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: *NIPS*, pp. 2843–2851 (2012)
6. Elhamifar, E., Sapiro, G., Sastry, S.: Dissimilarity-based sparse subset selection. *TPAMI* **PP**(99), 1 (2016)
7. Grady, L., Schwartz, E.L.: Isoperimetric graph partitioning for image segmentation. *TPAMI* **28**(1), 469–475 (2006)
8. Gretton, A., Borgwardt, K.M., Rasch, M., Schölkopf, B., Smola, A.J.: A kernel method for the two-sample-problem. In: *NIPS*, pp. 513–520 (2007)
9. Konukoglu, E., Glocker, B., Criminisi, A., Pohl, K.M.: Weighted spectral distance for measuring shape dissimilarity. *TPAMI* **35**(9), 2284–2297 (2013)
10. Liu, B., Huang, J., Kulikowski, C., Yang, L.: Robust visual tracking using local sparse appearance model and k-selection. *TPAMI* **35**(12), 2968–2981 (2013)
11. Long, M., Ding, G., Wang, J., Sun, J., Guo, Y., Yu, P.S.: Transfer sparse coding for robust image representation. In: *CVPR*, pp. 407–414 (2013)
12. Mori, G.: Guiding model search using segmentation. In: *ICCV*, vol. 2, pp. 1417–1423 (2005)
13. Qi, X., Xing, F., Foran, D.J., Yang, L.: Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set. *TBME* **59**(3), 754–765 (2012)
14. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: *CVPR*, pp. 3360–3367 (2010)
15. Xing, F., Xie, Y., Yang, L.: An automatic learning-based framework for robust nucleus segmentation. *TMI* **35**(2), 550–566 (2016)
16. Xing, F., Yang, L.: Fast cell segmentation using scalable sparse manifold learning and affine transform-approximated active contour. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 332–339. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-24574-4_40](https://doi.org/10.1007/978-3-319-24574-4_40)
17. Yu, K., Zhang, T., Gong, Y.: Nonlinear learning using local coordinate coding. In: *NIPS*, pp. 1–9 (2009)
18. Zhang, S., Zhan, Y., Metaxas, D.N.: Deformable segmentation via sparse representation and dictionary learning. *MedIA* **16**(7), 1385–1396 (2012)
19. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Deformable segmentation via sparse shape representation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011*. LNCS, vol. 6892, pp. 451–458. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-23629-7_55](https://doi.org/10.1007/978-3-642-23629-7_55)
20. Zimmer, C., Olivo-Marin, J.C.: Coupled parametric active contours. *TPAMI* **27**(11), 1838–1842 (2005)