Shape Prior Modeling using Sparse Representation and Online Dictionary Learning

Shaoting Zhang[1*], Yiqiang Zhan[2*],
Yan Zhou[3], Mustafa Uzunbas[2], Dimitris N. Metaxas[1]

[1]Department of Computer Science, Rutgers University, Piscataway, NJ, USA
[2]CAD R&D, Siemens Healthcare, Malvern, PA, USA
[3]Elekta Inc., Maryland Heights, MO, USA
* *Corresponding authors: shaoting@cs.rutgers.edu, yiqiang.zhan@siemens.com*

Abstract. The recently proposed Sparse Shape Composition (SSC) opens a new avenue for shape prior modeling. Instead of assuming any parametric model of shape statistics, SSC incorporates shape priors on-the-fly by approximating a shape instance (usually derived from appearance cues) by a sparse combination of shapes in a training repository. Theoretically, one can increase the modeling capability of SSC by including as many training shapes in the repository. However, this strategy confronts two limitations in practice. *First*, since SSC involves an iterative sparse optimization at run-time, the more shape instances contained in the repository, the less run-time efficiency SSC has. Therefore, a compact and informative shape dictionary is preferred to a large shape repository. *Second*, in medical imaging applications, training shapes seldom come in one batch. It is very time consuming and sometimes infeasible to reconstruct the shape dictionary every time new training shapes appear. In this paper, we propose an online learning method to address these two limitations. Our method starts from constructing an initial shape dictionary using the K-SVD algorithm. When new training shapes come, instead of re-constructing the dictionary from the ground up, we update the existing one using a block-coordinates descent approach. Using the dynamically updated dictionary, sparse shape composition can be gracefully scaled up to model shape priors from a large number of training shapes without sacrificing run-time efficiency. Our method is validated on lung localization in X-Ray and cardiac segmentation in MRI time series. Compared to the original SSC, it shows comparable performance while being significantly more efficient.

1 Introduction

Sparse Shape Composition (SSC) [12] is a recently proposed method for shape prior modeling. Different from previous methods [3], which often assume a parametric model for shape statistics, SSC is a non-parametric method that approximates an input shape usually derived from low level appearance features, by a sparse combination of other shapes in a repository. In this way, shape priors are incorporated on-the-fly. SSC is able to correct gross errors of input shape

and can preserve shape details even if they are not statistically significant in the training repository.

Theoretically, the more shape instances contained in the shape repository, the more shape modeling capacity SSC has. However, a repository including a large number of shapes adversely affects the efficiency of SSC, which iteratively performs sparse optimization at run-time. To reduce the redundancy in the shape repository and the computational cost, one can learn a compact and informative dictionary. Unfortunately, dictionary learning sometimes confronts another limitation. In medical imaging applications, training shape instances seldom come in one batch. If the dictionary needs to be completely re-learned every time when new training shapes come, the learning process will become very time consuming and sometimes infeasible.

In this paper, we propose an online learning method to address these two limitations. Our method starts from learning an initial dictionary offline using available training shapes. The K-SVD method is employed to learn the initial dictionary due to its flexibility and accelerated convergency. When new training shapes come, instead of re-constructing the dictionary from the ground up, we use an online dictionary learning method [8] to update the shape dictionary on-the-fly. With more and more new training shapes, our shape dictionary keeps updated to contain shape priors from all of them. Hence, sparse shape composition performed on the shape dictionary achieves two advantages: 1) The run-time efficiency of the shape composition is not sacrificed given much more training shapes. 2) SSC can be gracefully scaled-up to apply shape priors from, theoretically, infinite number of training shapes.

Relevant work: Related studies can be traced to two categories, shape modeling and sparse dictionary learning. In the former category, most previous studies [3, 6, 7, 10, 11] aim to model shape priors using a parametric model, e.g., multi-variant Gaussian [3] and hierarchical diffusion wavelet [7]. SSC is the first shape modeling method using sparse representation theory. Sparse dictionary learning methods have been extensively studied in signal processing domain. Popular ones include optimal direction (MOD) [5] and K-SVD [1]. While these methods require the access of all training samples, a recently proposed online dictionary learning [8] allows an efficient dictionary update only based on new samples. Although dictionary learning has been successfully applied on low level image processing tasks, to the best of our knowledge, the proposed method is the first one to employ them for high-level shape prior modeling.

2   Methodology

In this section, we will first briefly introduce standard Sparse Shape Composition. Dictionary learning technologies that aims to tackle the two limitations of SSC will be presented afterwards.

Sparse Shape Composition: SSC is designed based on two observations: 1) After being aligned to a common canonical space, any shape can be approximated by a sparse linear combination of other shape instances in the same shape

category. Approximation residuals might come from inter-subject variations. 2) If the shape to be approximated is derived by appearance cues, residual errors might include gross errors from detection/segementaion errors. However, such errors are sparse as well. Accordingly, shape priors can be incorporated *on-the-fly* through shape composition, which is formulated as a sparse optimization problem as follows.

In SSC, a shape is represented by a contour (2D) or a triangle mesh (3D) which consists of a set of vertices. Denote the input shape as $\mathbf{v}$, where $\mathbf{v} \in \mathbf{R}^{\mathfrak{D}N}$ is a vector concatenated by coordinates of its N vertices, where $\mathfrak{D} = \{2,3\}$ denotes the dimensionality of the shape modeling problem. (In the remainder of this paper, any shape instance is defined as a vector in the same way.) Assume $D = [\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_K] \in \mathbf{R}^{\mathfrak{D}N \times K}$ is a large shape repository that includes K accurately annotated and pre-aligned shape instances $\mathbf{d}_i$. The approximation of $\mathbf{v}$ by D is then formulated as an optimization problem:

$$\arg\min_{\mathbf{x},\mathbf{e},\beta} \|T(\mathbf{v},\beta) - D\mathbf{x} - \mathbf{e}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{e}\|_1, \qquad (1)$$

where $T(\mathbf{v}, \beta)$ is a global transformation operator with parameter β, which aligns the input shape $\mathbf{v}$ to the common canonical space of D. The key idea of SSC lies in the second and third terms of the objective function. In the second term, the $L1$-norm of $\mathbf{x}$ ensures that the nonzero elements in $\mathbf{x}$, i.e., the linear combination coefficients, is sparse [2]. Hence, only a sparse set of shape instances can be used to approximate the input shape, which prevents the overfitting to errors from missing/misleading appearance cues. In the third term, the same sparse constraint applies on $\mathbf{e} \in \mathbf{R}^{\mathfrak{D}N}$, the large residual errors, which incorporates the observation that gross errors might exist but are occasional. Eq. 1 is optimized using an Expectation-Maximization (EM) style algorithm, which alternatively optimizes β ("E" step) and $\mathbf{x}$, $\mathbf{e}$ ("M" step). "M" step employs a typical convex solver, e.g., interior-point convex solver [9] in this study.

Shape Dictionary Learning: Theoretically, the more shape instances in D, the larger shape modeling capacity SSC has. However, the run-time efficiency of SSC is also determined by the size of the shape repository matrix $D \in \mathbf{R}^{\mathfrak{D}N \times K}$. More specifically, the computational complexity of the interior-point convex optimization solver is $\mathcal{O}(N^2K)$ per iteration [9], which means the computational cost will increase quickly with the increase of K, the number of the shape instances in the shape repository. Note that $\mathcal{O}(N^2K)$ is the computational complexity for *one* iteration. Empirically, with larger K, it usually takes more iterations to convergency, which further decreases the algorithm speed.

In fact, owing to the similar shape characteristics across the population, these K shape instances usually contain lots of redundant information. Instead of including all of them, D should only contain "representative" shapes. This is exactly a dictionary learning problem, which has been extensively investigated in signal processing community. More specifically, a well learned dictionary should have a compact set of "atoms" that are able to sparsely approximate other signals. In our study, shape dictionary is learned using K-SVD [1], a popular dictionary learning method because of its accelerated converging speed.

Algorithm 1 Online learn and update dictionary, using mini-batch mode.

Input: Initialized dictionary $D_0 \in \mathbf{R}^{n \times k}$, input data $Y = [y_1, y_2, ..., y_K], y_i \in \mathbf{R}^n$, number of iterations T, regularization parameter $\lambda \in R$.
Output: Learned dictionary D_T.
$A_0 = 0, B_0 = 0$.
for $t = 1 \rightarrow T$ **do**
    Randomly draw a set of $y_{t,1}, y_{t,2}, ..., y_{t,\eta}$.
    for $i = 1 \rightarrow \eta$ **do**
        Sparse coding: $x_{t,i} = \arg\min_{x \in \mathbf{R}^k} \frac{1}{2}\|y_{t,i} - D_{t-1}x\|_2^2 + \lambda\|x\|_1$.
    end for
    $A_t = \beta A_{t-1} + \sum_{i=1}^{\eta} x_{t,i}x_{t,i}^T, B_t = \beta B_{t-1} + \sum_{i=1}^{\eta} y_{t,i}x_{t,i}^T,$
    where $\beta = \frac{\theta+1-\eta}{\theta+1}$, and $\theta = t\eta$ if $t < \eta$, $\theta = \eta^2 + t - \eta$ otherwise.
    Dictionary update: Compute D_t, so that:
    $\arg\min_{D} \frac{1}{t}\sum_{i=1}^{t} \frac{1}{2}\|y_i - Dx_i\|_2^2 + \lambda\|x_i\|_1 = \arg\min_{D} \frac{1}{t}\left(\frac{1}{2}Tr\left(D^T D A_t\right) - Tr(D^T B_t)\right).$
end for

Online Shape Dictionary Update: Using the compact dictionary derived by K-SVD, the run-time efficiency of SSC is dramatically improved, as the number of atoms in D is much less than the number of training shapes. However, K-SVD requires all training shapes available in the "dictionary update" step, which can not be satisfied in a lot of medical applications. For example, owing to the expensive cost, manual annotations of anatomical structures often come gradually from different radiologists/technicions. Re-construction of the dictionary D with every batch of new training shapes is very time consuming and not always feasible. To tackle this problem, we employ a recently proposed online dictionary method [8] to update the shape dictionary.

Algorithm 1 shows the framework of online dictionary learning for sparse coding. Starting from an initial dictionary learned by K-SVD, it iteratively employs two stages until converge, sparse coding and dictionary update. Sparse coding aims to find the sparse coefficient x_i for each signal y_i:

$$x_i = \arg\min_{x \in \mathbf{R}^k} \frac{1}{2}\|y_i - Dx\|_2^2 + \lambda\|x\|_1 \tag{2}$$

where D is the initialized dictionary or dictionary computed from the previous iteration. LARS-Lasso algorithm [4] is employed to solve this step. The dictionary update stage aims to update D based on all discovered $x_i, i \in [1, K]$:

$$\arg\min_{D} \frac{1}{K}\sum_{i=1}^{K} \frac{1}{2}\|y_i - Dx_i\|_2^2 + \lambda\|x_i\|_1 \tag{3}$$

Based on stochastic approximation, the dictionary is updated efficiently using block-coordinates descent. It is a parameter-free method and does not require any learning rate tuning. It is important to note that the "dictionary update" step in Algorithm 1 is significantly different from that of K-SVD. Instead of
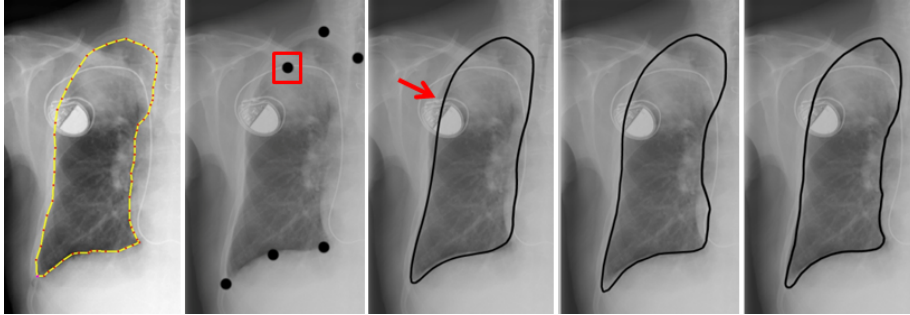
Fig. 1. Comparisons of the localization results. From left to right: manual label, detection results, PCA, SSC, and the online learning based shape refinement results. Due to the erroneous detection (marked by the red box), PCA result moves to the right and is not on the boundary (see the red arrow). Zoom in for better view.

requiring all training shapes, it only exploits a small batch of newly coming data (i.e., $x_i, i \in [1, \eta]$). The dictionary update thereby becomes much faster than K-SVD, as $\eta \ll K$. In this way, we can efficiently update the shape dictionary online by using new data as selected x_i.

Using this online updated dictionary, SSC obtains two additional advantages. 1) The run-time efficiency of shape composition is not sacrificed with much more training shapes. 2) SSC can be gracefully scaled-up to contain shape priors from, theoretically, infinite number of training shapes.

3   Experiments

We validate our algorithm in two applications, lung localization in Chest X-ray, and left ventricle tracking in MRI.

Lung localization: Chest radiography (X-ray) is a widely used medical imaging modality because of the fast imaging speed and low cost. Localization of lungs in chest radiography not only provides lung shapes, which are critical clues for pathology detection, but also paves the way for other medical image analysis tasks, e.g., cardiac measurements. On one hand, owing to the relatively cheap cost of manual/semi-automatic annotations of lungs in X-ray images, it is possible to get a large number of lung shapes for training. On the other hand, however, training lung shapes seldom come in one batch in clinical practices. Instead, clinicians often verify and correct auto-localization results and prefer a system that has self-improvement ability using these corrected shapes as new training shapes.Therefore, lung localization in chest X-ray becomes an ideal use case to test the effectiveness of our online dictionary method.

Our lung localization system starts from a set of auto-detected landmarks around the lung (e.g., the bottom-left lung tip), based on which lung shapes are inferred using shape priors. Note that various factors, e.g., imaging artifacts, lung diseases, etc., might induce missing/wrong landmark detection, which should be

corrected by shape prior models. Although the overall system performance depends on multiple components, including initial landmark detection, shape prior modeling and the following deformable segmentation, our comparison focuses on the shape prior modeling part, i.e., other components remain the same in comparsions. Our experimental dataset includes 367 X-ray images from different patients. 32 of them are used as training data to construct the initial data matrix/dictionary D in Eq. 1. Note that simply stacking more training shapes into D can also improve the capability of shape representation. However, it dramatically reduces the computational efficiency, which highly depends on the scale of D when solving Eq. 1 [9].

Three shape prior methods are compared, 1) the PCA based prior as used in Active Shape Model [3], 2) SSC [12], and 3) our method. Fig. 1 shows an example of using these methods to infer shapes from auto-detected landmarks. This case is challenging due to the misplaced medical instrument, which causes erroneous detections (marked by a red box in Fig. 1). Although all three methods achieve reasonable accuracy, the whole shape of PCA result shifts slightly to the right (where the red arrow points in

	P	**Q**	**DSC**
PCA	87.5 ± 5.2	96.0 ± 3.1	90.1 ± 4.0
SSC	86.7 ± 4.8	$\mathbf{96.6 \pm 2.4}$	89.4 ± 3.9
Ours	$\mathbf{94.3 \pm 4.6}$	96.2 ± 2.3	$\mathbf{94.5 \pm 3.6}$

Table 1. Quantitative comparisons of the lung localization using shape priors. P, Q, DSC stand for the sensitivity, specificity, and dice similarity coefficient (%), respectively.
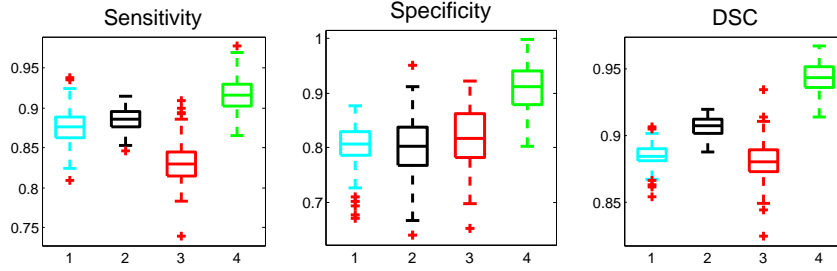
Fig. 1), because PCA is sensitive to outliers. Benefited by the sparse representation and $L1$-norm constraint, SSC and our method can both handle erroneous detections. However, since the initial shape dictionary may not be generative and representative enough, the inferred shape from SSC is not as accurate as the proposed method, which updates the dictionary on-the-fly and improves its capability of shape representations. Table 1 shows the quantitative accuracy (compared to experts' annotations) of the three methods, in terms of the sensitivity, specificity, and dice similarity coefficient (DSC). In general, our method achieves significantly better sensitivity, while slightly worse specificity than SSC. The reason is that SSC under-segments some images, which results in low sensitivity but high specificity. Our method achieves much better performance in terms of DSC, which is a more comprehensive measurements (includes both sensitivity and specificity) for localization accuracy. The experiments are performed on a PC with 2.4GHz Intel Quad CPU, 8GB memory, with Python 2.5 and C++ implementations. The whole framework is fully automatic and efficient. The shape inference step takes 0.2-0.3s, with around 0.06s as an overhead to update the dictionary online, which is negligible. In contrast, re-training the dictionary using K-SVD needs around 15-40s each time.

Real-time left ventricle tracking: Extraction of the boundary contour of a beating heart from cardiac MRI image sequences plays an important role

Fig. 2. Box plots for quantitative comparisons. Blue, black, red and green boxes represent results from the deformation, PCA, SSC, and the proposed method, respectively.

in cardiac disease diagnosis and treatments. MRI-guided robotic intervention is potentially important in cardiac procedures such as aortic valve repair. One of the major difficulties is the path planning of the robotic needle, which requires accurate contour segmentation of the left ventricle on a *real-time* MRI sequence. Thus, the algorithm should be robust, accurate and fast. We use a shape prior based tracking framework to solve this problem. In our method, a collaborative trackers network is employed to provide a deformed mesh and then generate a rough contour as the initialization at each time step [13]. Next, this initialized shape model deforms based on low level image appearance. Appearance-based deformation may not be accurate since the image information can be ambiguous and noisy. Thus, the shape prior model is employed to refine the deformed contour. Based on this framework, we compare the performance of (a) deformable model based on image appearance, (b) PCA based, (c) SSC based and (d) the online dictionary based shape refinement methods. For computational efficiency consideration, the dictionary size of (b) and (c) is fixed as a small number 8. The SSC method constantly uses this initial dictionary, while the proposed method (c) updates the dictionary on-the-fly by using acquired tracking results as the mini-batch input of Algorithm 1.

Fig. 2 shows the quantitative evaluations, in terms of the sensitivity, specificity, and the dice similarity coefficient. Appearance-based deformation results produces inconsistent results when the image information is ambiguous. SSC based shape refinement may not improve the accuracy of the deformed result due to the small size of dictionary. PCA based method achieves good performance. However, it is not able to handle certain new shapes which cannot be generalized from the current PCA results. In general, the proposed method achieves the most accurate result, since it updates the dictionary on-the-fly using newly acquired information. Thus it is more generic and adaptive to new data. Online updating the dictionary takes around 0.03s, which causes very small overhead for the whole system. To track total of 189 frames, our system takes 23.7s. Retraining the dictionary using K-SVD takes around 12s each time, which is not feasible for realtime applications.

4   Conclusions

In this paper, we propose a shape modeling method to tackle the two limitations of Sparse Shape Composition (SSC). Instead of directly including all training shapes in a repository, we employ dictionary learning technologies to learn a compact and informative shape dictionary. In more details, an initial shape dictionary is learned by K-SVD using available training shapes. When new training shapes come, online dictionary learning method is used to update the dictionary on-the-fly. With the dynamic updated dictionary, SSC is gracefully scaled-up to contain shape priors from a large number of training shapes without losing the run-time efficiency. Compared to standard SSC, it achieved better shape modeling performance with a much faster speed.

References

1. Aharon, M., Elad, M., , Bruckstein, A.: K-SVD: An algorithm for designing over-complete dictionaries for sparse representation. In: TSP. pp. 4311–4322 (2006)
2. Candes, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory 52(2), 489–509 (2006)
3. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape model - their training and application. Computer Vision and Image Understanding 61, 38–59 (1995)
4. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. The Annals of statistics 32(2), 407–499 (2004)
5. Engan, K., Aase, S., Hakon Husoy, J.: Method of optimal directions for frame design. In: ICASSP. vol. 5, pp. 2443–2446 (1999)
6. Hufnagel, H., Pennec, X., Ehrhardt, J., Handels, H., Ayache, N.: Shape analysis using a point-based statistical shape model built on correspondence probabilities. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I, LNCS, vol. 4791, pp. 959–967. Springer, Heidelberg (2007)
7. Langs, G., Paragios, N., Essafi, S.: Hierarchical 3D diffusion wavelet shape priors. In: ICCV. pp. 1717–1724 (2010)
8. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: ICML. pp. 689–696 (2009)
9. Nesterov, Y., Nemirovsky, A.: Interior point polynomial methods in convex programming. Studies in applied mathematics 13, 1993 (1994)
10. Shi, Y., Qi, F., Xue, Z., Chen, L., Ito, K., Matsuo, H., Shen, D.: Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics. IEEE Transactions on Medical Imaging 27(4), 481 –494 (2008)
11. Yan, P., Kruecker, J.: Incremental shape statistics learning for prostate tracking in trus. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 42–49. Springer, Heidelberg (2010)
12. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D., Zhou, X.: Towards robust and effective shape modeling: Sparse shape composition. Medical Image Analysis 16(1), 265 – 277 (2012)
13. Zhou, Y., Yeniaras, E., Tsiamyrtzis, P., Tsekos, N., Pavlidis, I.: Collaborative tracking for MRI-guided robotic intervention on the beating heart. In: Jiang, T., Navab, N., Pluim, J., Viergever, M. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 351–358. Springer, Heidelberg (2010)