DeepShape: Deep Learned Shape Descriptor for 3D Shape Matching and Retrieval

Jin Xie[1], Yi Fang[1], Fan Zhu[1], Edward Wong[2]
[1]New York University Abu Dhabi. [2]New York University Polytechnic School of Engineering.

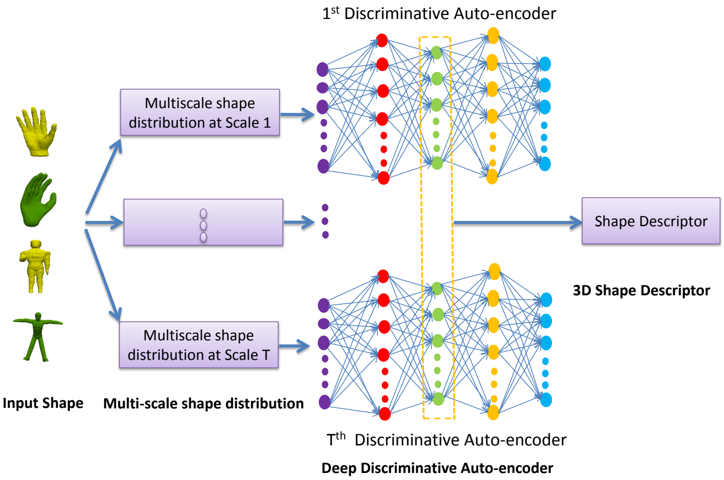

Figure 1: The framework of the proposed shape descriptor.

Complex geometric structural variations of 3D models usually pose great challenges in 3D shape matching and retrieval. In the past decades, plenty of shape descriptors have been proposed, such as the $D2$ shape distribution [4], statistical moments [7]. Apart from the earlier shape descriptors, another widely used shape signature is heat kernel signature (HKS) [6], where Sun *et al*. proposed to use the diagonal of the heat kernel as a local descriptor to represent shape. Nonetheless, these shape descriptors are hand-crafted rather than learned from a set of training shapes. In [2], the authors applied the bag-of-features (BOF) paradigm to learn the shape descriptor. In this paper, we develop a novel auto-encoder based shape descriptor for retrieval, which imposes the Fisher discrimination criterion on the hidden layer to make the hidden layer features discriminative and insensitive to geometric structure variations. It is expected that the neurons in the hidden layer have small within-class scatter but big between-class scatter. Moreover, we train multiple discriminative auto-encoders and concatenate all neurons in the hidden layers as the high-level learned shape descriptor for retrieval.

We detail the proposed framework of the discriminative auto-encoder based shape descriptor, which comprises three components, namely, multiscale shape distribution, discriminative auto-encoder and 3D shape descriptor. Figure 1 shows the proposed framework. In the multiscale shape distribution component, the distributions of heat kernel signatures of shape at different scales are extracted as a low-level feature for use as input to the discriminative auto-encoder. Then we train a discriminative auto-encoder to learn a high level feature embedded in the hidden layer of the discriminative auto-encoder component. In the 3D shape descriptor component, we form a descriptor from all hidden layer representations of the multiple discriminative auto-encoders.

Shape distribution [5] refers to a probability distribution sampled from a shape function describing the 3D model. We can consider HKS at each scale as a shape function defined on the surface of a 3D model. Then the shape distribution can be defined as the probability distribution of the shape function. In this work, we use histogram to estimate the probability distribution. For the scale t, we calculate the histogram of $\boldsymbol{S}^t_{i,j}$ of N vertices of the shape $\boldsymbol{y}_{i,j}$ to form the shape distribution $\boldsymbol{h}^t_{i,j}$. By considering probability distributions of shape functions derived from HKS at different scales, a multiscale shape distribution can be developed.

In order to boost the discriminative power of the shape distribution, based on the auto-encoder [1], we propose a discriminative auto-encoder for shape retrieval by imposing a Fisher discrimination criterion [3] on the hidden features. Based on the Fisher discriminative criterion, the discrimination can be achieved by minimizing the within-class scatter of $\boldsymbol{z}^t$, denoted by $S_w(\boldsymbol{z}^t)$, and maximizing the between-class scatter of $\boldsymbol{z}^t$, denoted by $S_b(\boldsymbol{z}^t)$. $S_w(\boldsymbol{z}^t)$ and $S_b(\boldsymbol{z}^t)$ are defined as:

$$
\begin{aligned}
S_w(\boldsymbol{z}^t) &= \sum_{i=1}^{C} \sum_{\boldsymbol{z}^t_{i,j} \in i} (\boldsymbol{z}^t_{i,j} - \boldsymbol{m}^t_i)(\boldsymbol{z}^t_{i,j} - \boldsymbol{m}^t_i)_i)^T \\
S_b(\boldsymbol{z}^t) &= \sum_{i=1}^{C} n_i (\boldsymbol{m}^t_i - \boldsymbol{m}^t)(\boldsymbol{m}^t_i - \boldsymbol{m}^t)^T
\end{aligned} \tag{1}
$$

where $\boldsymbol{m}^t_i$ and $\boldsymbol{m}^t$ are the mean vector of $\boldsymbol{z}^t_i$ and $\boldsymbol{z}^t$, respectively, and n_i is the number of samples of class i. Intuitively, we can define the discriminative regularization term $L(\boldsymbol{z}^t)$ as $tr(S_w(\boldsymbol{z}^t)) - tr(S_b(\boldsymbol{z}^t))$. Thus, by incorporating the discriminative regularization term into the standard auto-encoder model, we can form the following objective function of the discriminative auto-encoder:

$$
\begin{aligned}
J(\boldsymbol{W}^t, \boldsymbol{b}^t) = argmin_{\boldsymbol{W}^t, \boldsymbol{b}^t} \sum_{i=1}^{C} \frac{1}{2} \|\boldsymbol{x}^t_i - G(F(\boldsymbol{x}^t_i))\|_2^2 \\
+ \frac{1}{2}\lambda \|\boldsymbol{W}^t\|_F^2 + \frac{1}{2}\gamma(tr(S_w(\boldsymbol{z}^t)) - tr(S_b(\boldsymbol{z}^t))).
\end{aligned} \tag{2}
$$

To optimize the objective function of the discriminative auto-encoder, we adopt the back-propagation method of the error. Implementation of the optimization is described in the paper.

In order to characterize the intrinsic structure of the shape more effectively, we train multiple discriminative auto-encoders by setting multiscale shape distributions to the inputs of the discriminative auto-encoder. That is, for each scale t, we can learn $\boldsymbol{W}^t$ and $\boldsymbol{b}^t$ from a set of training shape distributions, i.e., $\boldsymbol{x}^t_1, \boldsymbol{x}^t_2, \cdots, \boldsymbol{x}^t_C$, $t = 1, 2, \cdots, T$. Thus, T discriminative auto-encoders can be formed by T groups of shape distributions. Once the multiple discriminative auto-encoders are trained, we can concatenate the activations of all hidden layers to form a shape descriptor.

Our conclusion is that the proposed deep shape descriptor with the discriminative auto-encoder for shape matching and retrieval is insensitive to geometric structure variations and can achieve good performance in various tests for matching and retrieving 3D shapes.

[1] Yoshua Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009. doi: 10.1561/2200000006. URL http://dx.doi.org/10.1561/2200000006.

[2] Alexander M. Bronstein, Michael M. Bronstein, Leonidas J. Guibas, and Maks Ovsjanikov. Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Transactions on Graphics*, 30(1):1, 2011. doi: 10.1145/1899404.1899405. URL http://doi.acm.org/10.1145/1899404.1899405.

[3] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Ed)*. Wiley, 2001.

[4] Michael Elad, Ayellet Tal, and Sigal Ar. Content based retrieval of VRML objects - an iterative and interactive approach. *Proceedings of the 6th Eurographics Workshop Multimedia*, pages 97–108, 2001.

[5] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dokin. Shape distributions. *ACM Transactions on Grphics*, 33:133–154, 2002.

[6] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Proceedings of the Symposium on Geometry Processing*, pages 1383–1392, 2009.

[7] D. V. Vranic, D. Saupe, and J. Richter. Tools for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics. *IEEE MMSP*, pages 293–298, 2001.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.