SparseDFF: Sparse-View Feature Distillation for One-Shot Dexterous Manipulation

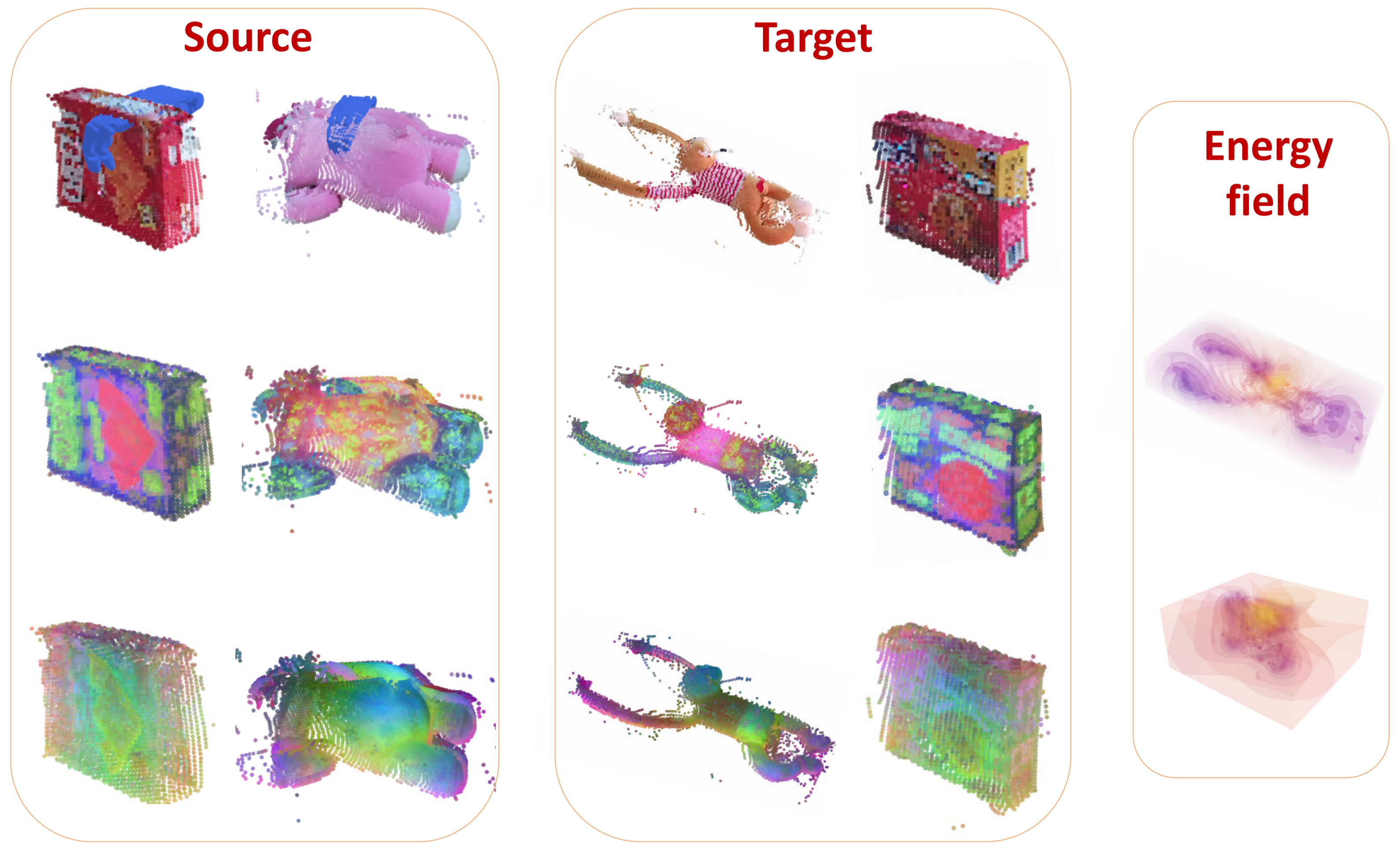
Qianxu Wang[1,3] Haotong Zhang[1] Congyue Deng[2,✉] Yang You[2] Hao Dong[1] Yixin Zhu[3,4,✉] Leonidas Guibas[2,✉]

1 CFCS, School of Computer Science, Peking University, China
2 Department of Computer Science, Stanford University, USA
3 Institute for AI, Peking University, China
4 PKU-WUHAN Institute for Artificial Intelligence, China

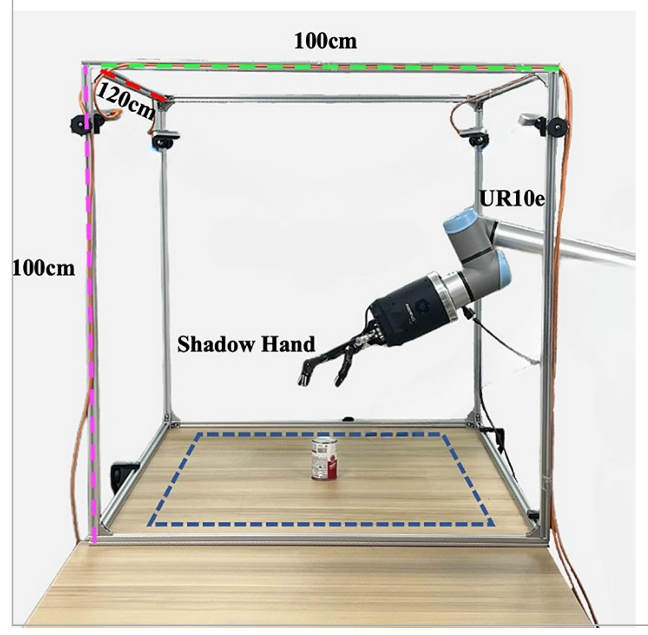
Introduction

Overview of SparseDFF

We introduce a novel method, SparseDFF, for distilling view-consistent **3D Distilled Feature Field (DFF)** from sparse RGBD images, readily generalizable to novel scenes without any modifications or fine-tuning. The DFFs create dense correspondences across scenes, enabling **one-shot** learning of **dexterous manipulations**. This approach facilitates seamless manipulation transfer to new scenes, effectively handling variations in object poses, deformations, scene contexts, and categories.



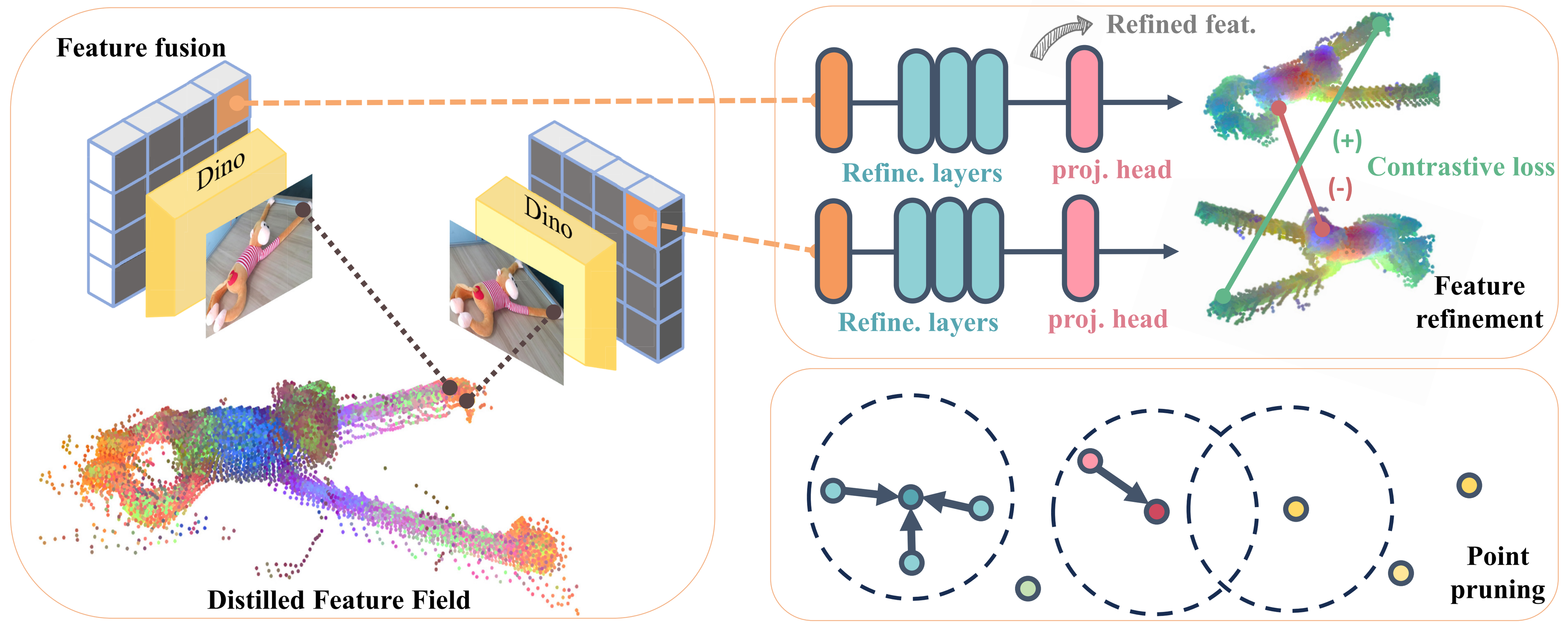
Our approach can be applied to various objects, bears, boxes, monkeys, and so on to convert an object, which can be a source or a target, to a feature field then optimized the feature field to the one with higher consistency.



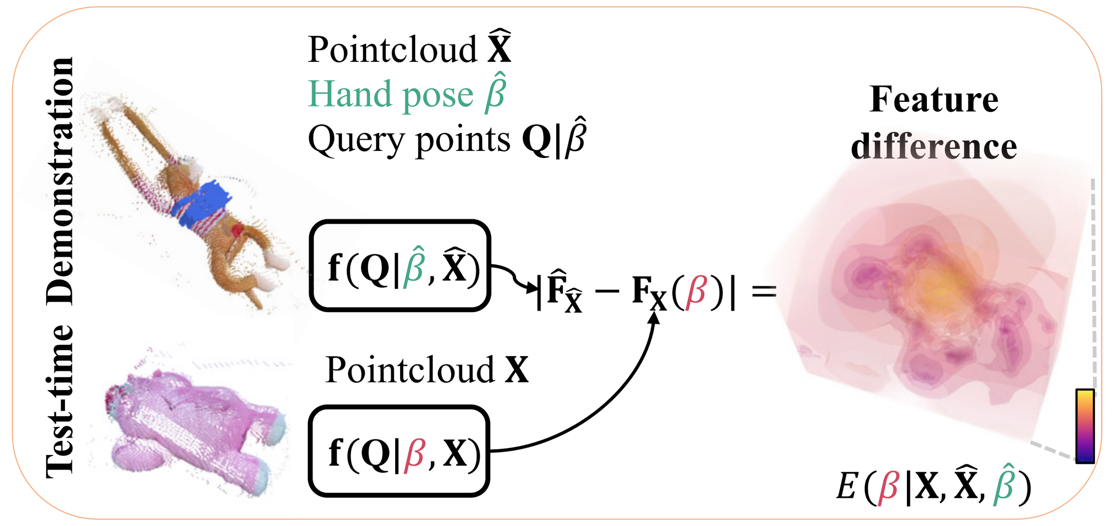
Our real robot setup:

Our methodology is validated through **real-world experiments** with a dexterous hand interacting with both rigid and deformable objects Our real robot setup consists of four Kinect cameras hanging on four pillars used to get point clouds of the scene, a UR 10e arm, a dexterous hand, and an object on the table to be manipulated.

Method



Constructing sparse-view DFFs Starting with the aggregation of DINO features, we form an initial 3D DFF. Next, a lightweight network then refines these features, trained solely on a single demonstration and employing contrastive loss to improve field consistency. Finally, a pruning algorithm assesses points through feature similarity in their vicinity. Points with minimal votes are eliminated.



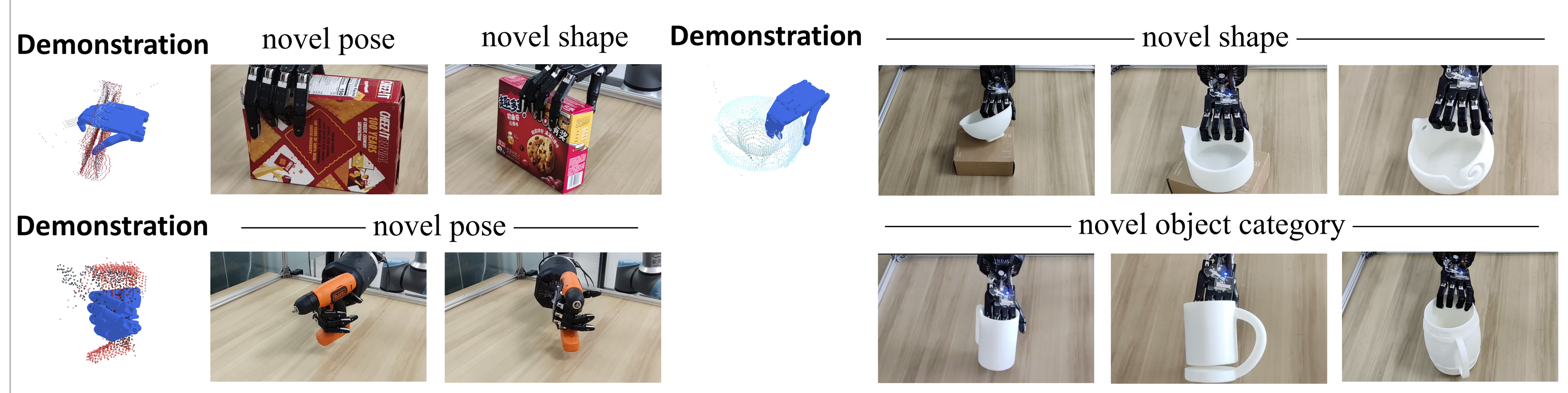
Optimization objective: We sample query points on the end-effector and compute their features using the learned 3D feature field. Minimizing the feature differences as an **energy function** facilitates the transfer of the end-effector pose from the source demonstration to the target manipulation.

Optimization process: The color gradient on the hand indicates the optimization steps from start to end.

Energy Function:

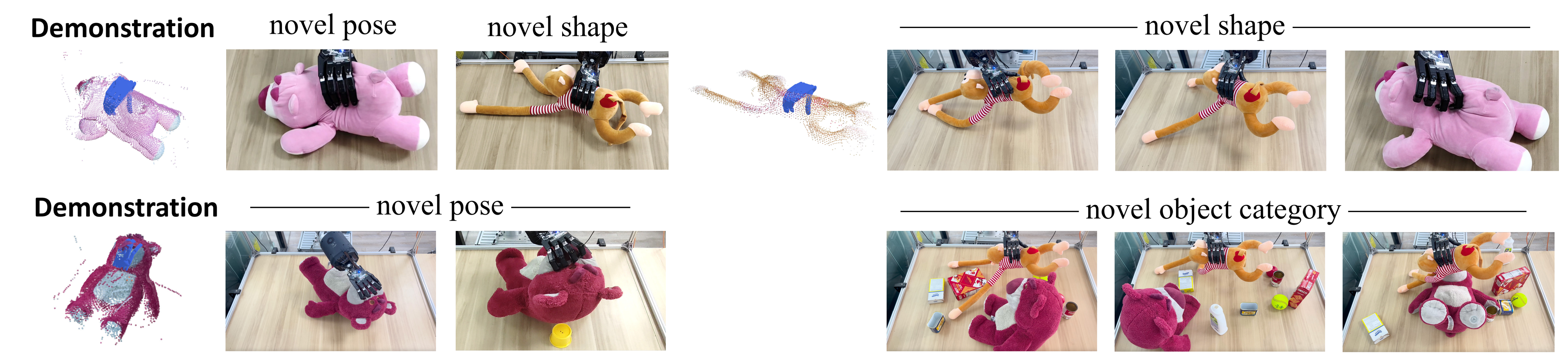
$$E(\beta|\mathbf{X},\hat{\mathbf{X}},\hat{\beta}) = |\mathbf{f}(\mathbf{Q}|\hat{\beta},\hat{\mathbf{X}}) - \mathbf{f}(\mathbf{Q}|\beta,\mathbf{X})|.$$

Experiments



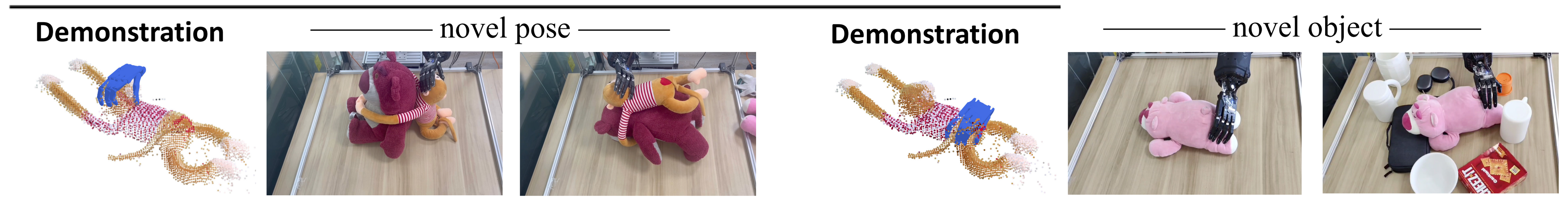
Qualitative results on rigid objects grasping: Each panel illustrates the initial grasping pose, determined via our end-effector optimization, followed by a frame capturing the successful lift-off of the target object. Grasping Box1 and transferring the skill to Boxes in new poses, including a distinct box Box2. A functional grasp of a drill by its handle. Transferring the learned grasp on Bowl1 to bowls with varied shapes (top row) and cross-category generalization to Mugs (bottom row).

Demo.	Box1		Drill	Bowl1					
Target	Box1	Box2	Drill	Bowl1	Bowl2	CatBowl	Mug	FloatingMug	BeerBarrel
UniDexGrasp++*	7.7%	-	66.9%	37.7%	31.9%	26.3%	24.7%	25.5%	6.2%
DFF	90%	0%	**100%**	**100%**	0%	30%	0%	20%	10%
Ours	**100%**	**100%**	**100%**	**100%**	**80%**	**60%**	**80%**	**40%**	**90%**



Qualitative results on deformable objects grasping: For each successful grasp, we show the initial grasping pose and a frame demonstrating the successful lift of the object off the table. Learning to grasp SmallBear and transferring this skill to various nose. Learning to grasp the Monkey, showcasing adaptability to significant deformations and transfers to SmallBear.

Demo.	Monkey			BigBear	SmallBear	
Target	Monkey	MonkeyScene	SmallBear	BigBear	SmallBear	Monkey
DFF	90%	40%	0%	20%	**90%**	0%
Ours	**100%**	**100%**	**60%**	**80%**	**90%**	**50%**



Pet toy animals: Head caressing and butt patting is transferred from a single, lying Monkey to a scene with the Monkey hugging the BigBear, SmallBear, respectively.