

# Canonical Volume（规范体积）

---

Canonical volume（规范体积）在计算机视觉和图形学中通常指一个标准化的三维空间，用于对物体或场景进行表示和分析。它的主要特点是通过将物体或场景的几何和属性映射到一个固定的、通用的坐标系中，方便后续的计算和操作。

---

## Canonical Volume 的定义

---

- **规范性：** Canonical volume 通常是一个标准化的 3D 空间，比如：

$$[-1, 1] \times [-1, 1] \times [-1, 1] \quad (29)$$

其中所有的几何和属性都被映射到这个空间。

- **固定性：** 它不依赖具体的视角、场景或对象的姿态，是一个全局的“基准”。
  - **动态性（如果需要）：** 在动态场景中，Canonical volume 可以与时间相关，用于描述随时间变化的场景。
-

# Canonical Volume 的作用

---

## 1. 标准化表示：

- Canonical volume 通过将物体或场景的属性映射到一个标准坐标系中，消除了输入数据中可能存在的位姿、大小、视角等差异，便于后续的处理。
- 例如，在人体姿态估计中，可以将人体的不同姿势归一化到一个规范体积中，方便分析。

## 2. 3D 几何描述：

- Canonical volume 可以用来表示场景的 3D 几何结构，如点云、体素网格、隐式场等。

## 3. 作为计算的桥梁：

- 在很多任务中，Canonical volume 是一种“中间表示”，比如从 2D 图像提取特征后，将特征映射到 Canonical volume，再从 Canonical volume 映射回 2D 或其他视图。

---

# Canonical Volume 的特点

---

## 1. 不依赖视角：

- Canonical volume 中的表示是物体自身的固有属性，与摄像机的角度无关。这对于多视角融合或新视角生成尤为重要。

## 2. 便于映射和操作：

- 物体或场景可以通过映射函数：

$$f(x) \quad (30)$$

投影到 Canonical volume 中，也可以通过逆映射：

$$f^{-1}(x) \quad (31)$$

恢复到原空间。

## 3. 支持动态特性：

- 在动态场景中，Canonical volume 可以通过时间维度扩展为：

$$V(x, t) \quad (32)$$

以描述随时间变化的物体或场景。

---

# Canonical Volume 在视频高斯表示 (VGR) 中的应用

---

在 VGR 方法中，Canonical volume 的使用主要体现在以下方面：

### 1. 规范化表示：

- 视频帧中的场景内容（几何结构、运动轨迹等）被统一映射到 Canonical volume 中，通常使用正交相机坐标系（如文中所述的  $X$ 、 $Y$ 、 $Z$  轴分别对应宽度、高度、深度）。

### 2. 作为 3D 高斯的基准空间：

- 3D 高斯的位置、旋转、缩放等参数在 Canonical volume 中进行优化，形成时间无关的标准表示。

### 3. 与 2D 投影结合：

- Canonical volume 的表示通过投影函数：

$$\pi(x) \tag{33}$$

映射到 2D 图像平面，便于和 2D 光流、深度等先验信息结合。

---

## 实例

假设有一个视频帧序列，其中某个场景包含一个物体从左向右移动。以下是 Canonical volume 的作用：

## 1. 输入场景：

- 输入的场景是一个动态 3D 场景，在每帧图像中，物体的位置随时间变化。

## 2. 映射到 Canonical volume：

- 将每帧图像的 3D 内容映射到 Canonical volume 中，比如将物体的中心标准化到：

$$(0, 0, 0) \quad (34)$$

移动路径在 Canonical volume 中用统一的轨迹表示。

## 3. 优化与处理：

- 在 Canonical volume 中优化物体的位置、运动轨迹和外观信息，使得这些参数独立于场景中的视角。

## 4. 渲染输出：

- 通过逆映射函数：

$$f^{-1}(x) \quad (35)$$

将 Canonical volume 中的优化结果映射回原始视频的每个帧，从而生成一致性高、可编辑的重建结果。

---

# 总结

---

Canonical volume 是一个标准化的 3D 空间，广泛用于视频表示、三维重建和动态场景的建模中。通过将内容映射到 Canonical volume，可以显著简化多视图融合、动态建模等任务的难度，同时增强结果的鲁棒性和一致性。