AANet: Adaptive Aggregation Network for Efficient Stereo Matching

Supplementary Material

Haofei Xu  Juyong Zhang*
University of Science and Technology of China
xhf@mail.ustc.edu.cn, juyong@ustc.edu.cn

In this supplementary document, we briefly review traditional cross-scale cost aggregation algorithm [1] to make this paper self-contained.

For cost volume $C \in \mathbb{R}^{D \times H \times W}$, [1] reformulates the local cost aggregation from an optimization perspective:

$$\hat{C}(d, p) = \arg \min_{z} \sum_{q \in N(p)} w(p, q) \| z - C(d, q) \|^2,$$  

(1)

where $\hat{C}(d, p)$ denotes the aggregated cost at pixel $p$ for disparity candidate $d$, pixel $q$ belongs to the neighbors $N(p)$ of $p$, and $w$ is the weighting function to measure the similarity of pixel $p$ and $q$. The solution of this weighted least square problem (1) is

$$\hat{C}(d, p) = \sum_{q \in N(p)} w(p, q) C(d, q).$$  

(2)

Thus, different local cost aggregation methods can be reformulated within a unified framework.

Without considering multi-scale interactions, the multi-scale version of Eq. (1) can be expressed as

$$\hat{v} = \arg \min_{\{z^s\}_{s=1}^{S}} \sum_{s=1}^{S} \sum_{q^s \in N(p^s)} w(p^s, q^s) \| z^s - C^s(d^s, q^s) \|^2,$$  

(3)

where $p^s$ and $d^s$ denote pixel and disparity at scale $s$, respectively, and $p^{s+1} = p^s / 2, d^{s+1} = d^s / 2, p^1 = p$ and $d^1 = d$. The aggregated cost at each scale is denoted as

$$\hat{v} = [\hat{C}^1(d^1, p^1), \hat{C}^2(d^2, p^2), \cdots, \hat{C}^S(d^S, p^S)]^T.$$  

(4)

The solution of Eq. (3) is obtained by performing cost aggregation at each scale independently:

$$\hat{C}^s(d^s, p^s) = \sum_{q^s \in N(p^s)} w(p^s, q^s) C^s(d^s, q^s),$$  

(5)

By enforcing the inter-scale consistency on the cost volume, we can obtain the following optimization problem:

$$\hat{v} = \arg \min_{\{z^s\}_{s=1}^{S}} \left( \sum_{s=1}^{S} \sum_{q^s \in N(p^s)} w(p^s, q^s) \| z^s - C^s(d^s, q^s) \|^2 + \lambda \sum_{s=2}^{S} \| z^s - z^{s-1} \|^2 \right),$$  

(6)

where $\lambda$ is a parameter to control the regularization strength, and $\hat{v}$ is denoted as

$$\hat{v} = [\hat{C}^1(d^1, p^1), \hat{C}^2(d^2, p^2), \cdots, \hat{C}^S(d^S, p^S)]^T.$$  

(7)

The optimization problem (6) is convex and can be solved analytically (see details in [1]). The solution can be expressed as

$$\hat{v} = P\hat{v},$$  

where $P$ is an $S \times S$ matrix. That is, the final cost volume is obtained through the adaptive combination of the results of cost aggregation performed at different scales.

Inspired by this conclusion, we design our cross-scale cost aggregation architecture as

$$\hat{C}^s = \sum_{k=1}^{S} f_k(\hat{C}^k), \quad s = 1, 2, \cdots, S,$$  

(9)

where $f_k$ is defined by neural network layers.

References


*Corresponding author