Learning About SGM-Nets II

Hongzhi Liu

May 9, 2018

1 SGM-Net

In the last article, I learned a lot about *SGM* (Semi-global matching) including its strengths and weaknesses [1]. Besides, I briefly introduced an overview of *SGM-Net*. Then I will continue to learn about standard parameterization of *SGM* and an architecture of *SGM-Net*.

According to the thesis of Professor Seki, a necessary condition to obtain the correct disparity is that a path traversing the correct disparity $d_{gt}^{x_0}$ at pixel x_0 should be smaller than any other paths, i.e. a cost L_r at pixel x_0 must satisfy $L_r\left(x_0, d_i^{x_0}\right) > L_r\left(x_0, d_{gt}^{x_0}\right)$, $\forall d_i \in [0, d_{max}] \neq d_{gt}$. Ant they formulate it with a hinge loss function as Eq. (1):

$$E_g = \sum_{d_i^{x_0} \neq d_{gt}^{x_0}} max\left(0, L_r\left(x_0, d_i^{x_0}\right) - L_r\left(x_0, d_{gt}^{x_0}\right) + m\right) \tag{1}$$

where m means margin. The hinge loss function allows easier formulation of back-propagation compared to other functions such as softmax loss. In order to allow the backpropagation of the loss function, they clarify the gradients of Eq. (1) with respect to p_1 and p_2. I can see an example in Fig. 1 .

$$\begin{aligned} \frac{\partial E_g}{\partial P_{1,r}} &= \sum_{d_i^{x_0} \neq d_{gt}^{x_0}} \sum_n \left(T\left[\left|\delta d^{x_n \leftarrow d_{gt}^{x_0}}\right| = 1\right]\right. \\ &\qquad \left. - T\left[\left|\delta d^{x_n \leftarrow d_t^{x_0}}\right| = 1\right]\right) \\ \frac{\partial E_g}{\partial P_{2,r}} &= \sum_{d_i^{x_0} \neq d_{gt}^{x_0}} \sum_n \left(T\left[\left|\delta d^{x_n \leftarrow d_{gt}^{x_0}}\right| > 1\right]\right. \\ &\qquad \left. - T\left[\left|\delta d^{x_n \leftarrow d_t^{x_0}}\right| > 1\right]\right) \end{aligned} \tag{2}$$

With the Eq. (2), the team is able to minimize the loss function by using the standard framework, i.e. forward and back propagation. Therefore they call this loss function "Path cos".

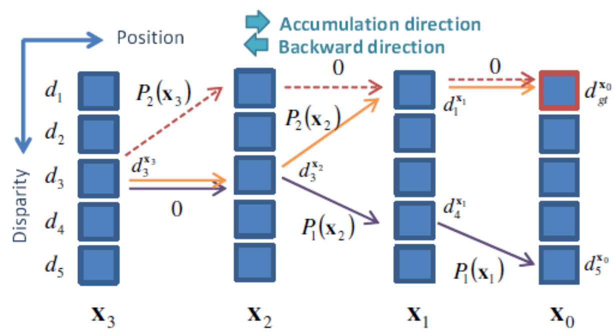

Figure 1: Consecutive 4 pixels and their 5 candidate disparities at each pixel. The orange and purple line represent the path fromcorrect disparity $d_{gt}^{x_0}$ and d_5 at root pixel x_0,respectively.

In order to remove the ambiguity of disparities traversed along the path, Professor Seki introduces "Neighbor cost" function. In order to compensate the advantage and difficulty of the path and neighbor costs, equations are put together and finally the loss function becomes in Eq. 3:

$$\begin{aligned} E = \sum_{r \in R} \Bigg(&\sum_{x_0, x_1 \in G_b} E_{n_b} + \sum_{x_0, x_1 \in G_s} E_{n_s} + \\ &\sum_{x_0, x_1 \in G_f} E_{n_f} + \xi \sum_{x_0 \in G} E_g \Bigg) \end{aligned} \tag{3}$$

where ξ means a blending ratio. They randomly extracted the same number of pixels for border G_b, slant

G_s, and flat G_f on each direction r. All $G*$ have annotation of true disparity. The disparity map given by *SGM-Nets* trained with Eq. (3) is shown in Fig. 2.

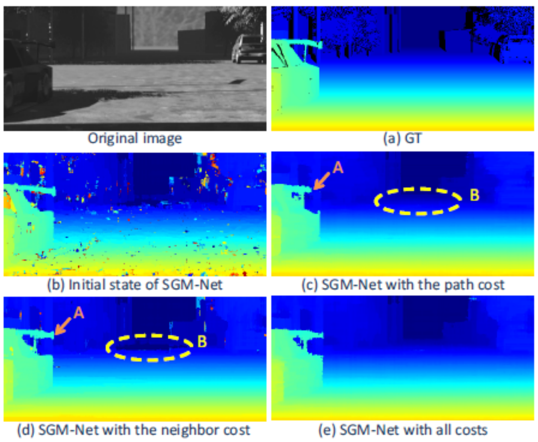

Figure 2: Comparison of the costs for the loss function.

2 Evaluation of SGM-Nets

Professor Seki evaluated *SGM-Nets* with the test images on the website. Table 1 shows estimated error and absolute ranking on K12. Signed and standard *SGM-Nets* got the 1st rank on K12.

Table 1: Out-Noc error on KITTI 2012 testing dataset by October 18th 2016. "*" means GPU computation.

Rank	Method	Error	Time [sec.]
1	Signed SGM-Net	2.29%	67*
2	Standard SGM-Net	2.33%	67*
3	PBCP [2]	2.36%	68*
4	Displets v2 [3]	2.37%	265
5	MC-CNN-acrt [4]	2.43%	67*

Besies, their method achieved the same accuracy on the overall criterion without the prior knowledge. Professor Seki emphasises that the most of computation time is consumed by stereo correspondence. *SGM-Nets* take only a few seconds on the GPU.

References

[1] Akihito Seki and Marc Pollefeys. Sgm-nets: Semi-global matching with neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6640–6649, 2017.

[2] Akihito Seki and Marc Pollefeys. Patch based confidence prediction for dense disparity map. In *British Machine Vision Conference*, pages 23.1–23.13, 2016.

[3] Fatma Guney and Andreas Geiger. Displets: Resolving stereo ambiguities using object knowledge. In *Computer Vision and Pattern Recognition*, pages 4165–4175, 2015.

[4] Yann Lecun. Stereo matching by training a convolutional neural network to compare image patches. 17(1):2287–2318, 2015.