Improved Stereo Matching based on Convolutional Neural Network

Yingnan Geng\textsuperscript{1,2}, Mi Lu\textsuperscript{2}, Baixiu Ding\textsuperscript{2,3}, Wenrui Zhang\textsuperscript{2}, and Yindong Hua\textsuperscript{2}

\textsuperscript{1}Electrical and Information Engineering, Changchun Institute of Technology, Changchun, Jilin, China
\textsuperscript{2}Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA
\textsuperscript{3}Electrical Engineering and Computer, Jilin Jianzhu University, Changchun, Jilin, China

\textbf{Abstract} — The MC-CNN (as Matching Cost Convolutional Neural Network) based on Convolutional Neural Network (CNN) obtains good results in stereo matching. However, it often takes tremendous time to run algorithms by GPU, because stereo images are big and the training module takes a lot of samples. In this paper, T-CNN, a stereo matching algorithm consisting of two main modules: training and testing is proposed. For the training module, fewer samples from the stereo image patches are trained in order to reduce the running time. For the testing module, we combine the values from the MC-CNN network, color information and gradient information to obtain cost matching function. In our experiments, the matching costs of T-CNN are compared with that of MC-CNN networks, showing that our algorithm T-CNN outperforms the MC-CNN.

\textbf{Keywords:} stereo matching, convolutional neural network, disparity map, color similarity, gradient similarity

\section{Introduction}

Stereo matching, finding corresponding pixels and obtaining disparity map from stereo images: reference image and target image, is always a basic problem in stereo vision and has been used in many fields such as 3D scene reconstruction, autonomous driving and intermediate view generation. For pixel \((x, y)\) in the reference image and pixel \((x - d, y)\) in the target image, the difference \(d\) in horizontal location of an pixel is called the disparity. Then the depth \(Z_{\text{depth}}\) can be computed by equation (1). And 3D coordinates \((X, Y, Z)\) can be obtained from (2), where \(f\) is the focal length of the camera and \(B\) is the distance between the camera centers. As depth and 3D coordinates can be obtained by disparity values, a stereo matching system aims to produce an accurate dense disparity map\textsuperscript{[1]}

\begin{equation}
Z_{\text{depth}} = \frac{fB}{d} \quad (1)
\end{equation}

\begin{equation}
\begin{aligned}
X &= \frac{Bx}{d} \\
Y &= \frac{By}{d} \\
Z &= \frac{Bf}{d}
\end{aligned} \quad (2)
\end{equation}

*This work was supported by the Science Foundation of education department of Jilin Province during ‘13rd Five-year plan (JKKH20191255KJ)."