Adaptive Rank Transform for Stereo Matching

Ge Zhao¹,², Yingkui Du¹, and Yandong Tang¹

¹ State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences
² Graduate University of the Chinese Academy of Sciences, Beijing, P.R. China
{zhaoge,dyk,ytang}@sia.cn

Abstract. Window selection is the main challenge for local stereo matching methods based on the rank transform and it involves two aspects: the rank window selection and the match window selection. Most recent methods only focus on how to select the match window but pay little attention to the selection of the rank window. In this paper, we propose a novel matching method based on adaptive rank transform. Differing with the existing rank-based matching methods, the proposed method can deal with the rank and match window selection at the same time. The experimental results are evaluated on the Middlebury dataset as well as real images, showing that our method performs better than the recent rank-based stereo matching methods.

Keywords: stereo matching, rank transform, window selection.

1 Introduction

Stereo matching [1] is the process of finding corresponding pixels between the left and right input images. It is widely used in many vision applications such as 3-D reconstruction. There are mainly two types of stereo matching algorithms which are global and local methods. Global methods [2,3,4,5] model the stereo matching problem as an energy minimization process. These methods usually have high matching accuracy and robustness but requires long running time and complex parameter tuning. In contrast, local stereo methods [6,7,8] are generally fast and easy to implement but less accurate than the global methods. To improve the matching accuracy, local methods commonly gather the information from all pixels in a window which is centered at the pixel under consideration and this window is called match window. Both the local and global methods face some common problems, which are the brightness differences, certain types of noise and radiometric distortions. One solution to these problems is to pre-process the stereo pair prior to matching. The rank transform [9] is one kind of these pre-processing methods, in which the intensity of the rank window center is replaced by the number of pixels in the rank window whose intensity value is more than the center pixel. Fig.1 gives an concrete example of the original rank transform and Fig.2 gives the basic flow of stereo matching algorithms based on the rank transform.
There are mainly two problems of stereo matching methods using the original rank transform or extensions of the original rank transform [10,11,12]. The first problem is the selection of the match window and this problem is well studied. For example, Zheng Gu and Xianyu Su [13] recently proposed a new stereo matching method using the rank transform. They assign adaptive weight to each pixel in a fixed match window according to the spatial proximity to the center of the match window. Kun Wang [14] also proposed a novel stereo matching algorithm using rank transform and he based his match window selection on edge detection. The second problem is the selection of an appropriate rank window for each pixel. However, until now, most rank-based methods pay little attention to this problem. In this paper, we are the first to deal with the rank and match window selection simultaneously.

The outline of this paper is as follows: in section 2, we propose the adaptive rank transform; in section 3, we propose a new way to construct self-adaptive match window based on adaptive rank windows; in section 4 we test our approaches with both benchmark and real stereo pairs and compare them with some recent matching methods; finally conclusion is summarized in section 5.

2 Adaptive Rank Transform

2.1 The Motivation Behind Adaptive Rank Transform

In stereo matching, window selection is a hard problem. In theory, the window should adapt to the local image structure which means only including the neighboring pixels from the similar depth of the scene with the pixel under consideration [15]. However, the 3D depth information is usually unknown beforehand and can only be estimated from the images. A common assumption is that pixels with similar intensities within a