Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

High-precision Human Body Acquisition via Multi-view Binocular Stereopsis <Supplementary Material>

1. Image Preprocessing

The 12 bit raw image format (i.e. CR2) is used to export images from cameras, and the background pixels are removed beforehand. When converting the raw data to float point RGB images, after de-Bayering we use various sets of parameter values of the photometric rendering operations, including the exposure, the contrast, and the highlights, to generate several images of the same viewpoint in rasterized image format (i.e., PNG). For instance, Fig. 1 shows three pairs of stereo images produced from the raw image data with different rendering parameters. In this way, the radiometric characteristics of the human body can be reserved as much as possible, which is beneficial to cost volume computation for depth recovery later on. Each pair of stereo images is first rectified to obtain row-aligned epipolar geometry [**?**], such that the matching progress can be conducted using mutual scanline. An example is shown in Fig. 2.

Fig. 1. (a),(b),(c) show three pairs of stereo images which are exported from the same stereo rig with three combinations of exposure, contrast and highlights. For example, the exposure value in (a) is larger than that in (b) and (c).

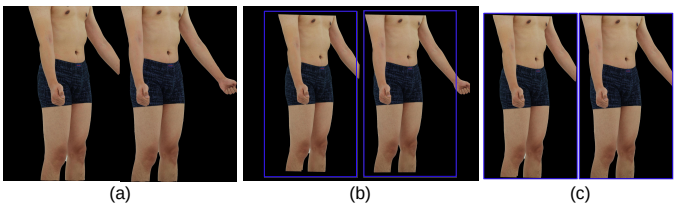

Fig. 2. Taking a pair of stereo image from Fig. 1(a) for example, the result of removing the background is shown in (a). (b) and (c) show the results with row-aligned geometry and cropped background, respectively.

2. Matching Cost

We use Zero-Normalized Cross Correlation (ZNCC) $C(p_r, p_m)$ as the matching cost when matching corresponding pixels. It is defined as:

$$\frac{\sum(I_r(q_r) - \bar{I}_r(p_r))(I_m(q_m) - \bar{I}_m(p_m))}{\sqrt{\sum(I_r(q_r) - \bar{I}_r(p_r))} \cdot \sqrt{\sum(I_m(q_m) - \bar{I}_m(p_m))}}, \tag{1}$$

where p_r, p_m are the matched pixels from the reference image I_r and the matching image I_m, respectively. By subtracting the mean intensity of a support area of p_r, p_m, indicated as $\bar{I}_r(p_r)$, $\bar{I}_m(p_m)$, we normalize the input images and take the adjacent structure into account to measure the matching similarity. Pixels in the support area of p_r, p_m are indicated as q_r,q_m, respectively.

3. Matching Constraints

Photometric Consistency. Given that matching ambiguities may be caused by texture similarity of human skin, this constraint aims to select the most reliable match from all candidates. It represents the distinctiveness of a match from its neighboring matches, and is defined as:

$$R(p_r, p_m) = \frac{C(p_r, p_m)}{max(C(p_r \pm 1, p_r \pm 1), \sigma)}, \tag{2}$$

where σ is a truncated parameter to prevent extremely large value. We use thresholds τ_c and τ_r for $C(p_m, p_r)$ and $R(p_m, p_r)$, to reject unreliable candidate matches. The larger τ_c and τ_r are, the more reliable matches will be obtained. In our experiments, $\tau_c = 0.95$ and $\tau_r = 1.5$ for extracting matching seeds, while the values are relaxed to 0.6 and 1.0 for seed-propagation.

Smoothness. This is to ensure similar disparity between a pixel and its neighbors. For a pixel p, we assume that more than half of the pixels in its neighborhood have similar disparity (difference is within 1):

$$\sum_{q \in N(p)} \mathrm{II}(|D(q) - D(p)| \leqslant 1) \geqslant \frac{|N(p)|}{2}, \tag{3}$$

where $N(p)$ denotes the neighborhood of p, and q is a pixel in $N(p)$. $\mathrm{II}(\cdot)$ is an indicator function. The best match should satisfy Eq. (3) in a 3×3 neighborhood.

Ordering. If an object point appears to be to the right of another object point in the left image, the relative position should be the same in the right image. Therefore, the disparity of p should not exceed the disparity of its adjacent pixels $p \pm 1$ by more than one.

Uniqueness. The stereo matching is performed commutatively. Two images can either be reference image and the matching image. A pixel p_m in the matching image I_m is matched with a pixel p_r in the reference image I_r, only if p_r is also matched with p_m the other way around.