Skin Color Weighted Disparity Competition for Hand Segmentation from Stereo Camera

Qi Wang[1]
wangqi@jdl.ac.cn

Xilin Chen[1]
xlchen@ict.ac.cn

Wen Gao[2]
wgao@pku.edu.cn

[1] Key Lab of Intelligent Information Processing, Institute of Computing Technology
Chinese Academy of Sciences
Beijing, China

[2] Institute of Digital Media, School of EE & CS
Peking University
Beijing, China

This paper focuses on realtime hand segmentation from stereo camera, where the simple thresholding operation with the dense disparity map is not appropriate since it is time consuming to compute the dense disparity map, and proposes a novel method of skin color weighted disparity competition. The basic idea is that the hand disparity will be apparently different from most of the background disparities, and thus the hand pixel can be validated by verifying whether it satisfies the disparity limit of the hand more than of the background. This is why we call our method *disparity competition*.

The framework of the proposed method is shown in Figure 1. It contains two main stages. The first stage marks the confident hand pixels and obtains all regions of interest (ROI) in the left image by incorporating skin color cue and depth cue from the detected face. The second stage validates the hand pixels for each ROI by employing skin color weighted disparity competition, as summarized as follows:

1. The skin color probability

$$\begin{cases} P(I) = \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(I-\mu_c)^t \Sigma^{-1}(I-\mu_c)\right] \\ \Sigma = \begin{pmatrix} \sigma_h^2 & & \\ & \sigma_s^2 & \\ & & \sigma_v^2 \end{pmatrix} \end{cases}, \quad (1)$$

where I is the probe pixel, μ_c is the mean HSV vector of all marked hand pixels in the current ROI, σ_h, σ_s or σ_v encodes respectively the standard deviation of the H, S or V component for the skin color in the current ROI.

2. The two competing disparity limits

$$\begin{cases} \chi^h = [u_d^h - e^h, u_d^h + e^h] \\ \chi^b(i,j) = [D^b(i,j) - e^b, D^b(i,j) + e^b] \end{cases}, \quad (2)$$

where the superscript of h or b represents the hand or background respectively, e^h or e^b encodes the maximal bias(typically set to 2 or 1 respectively), u_d^h is the mean disparity of all marked confident hand pixels in the current ROI and D^b is the result image of the WaterShed algorithm based on the reliable background pixels (resulted from the depth cur from the detected face) as the seed pixels and the zero image as the origin image.

3. The skin color weighted disparity competition

First introduce the disparity verification V as

$$V(L(i,j),d,N) = \frac{1}{2 \times N + 1} \sum_{n=-N}^{n=N} \|L(i,j+n), R(i,j+d+n)\|, \quad (3)$$

where $L(\cdot)$ or $R(\cdot)$ denotes the HSV color vector of the considering pixel in the left or right image respectively, d is the verifying disparity, N is the neighborhood size, and $\|\cdot\|$ represents the L_2 Norm;

and the skin probability of the local neighborhood of $I(i,j)$ as

$$\psi_I(i,j,N) \propto P\left(\frac{1}{2 \times N + 1} \sum_{n=-N}^{n=N} I(i,j+n)\right), I \in \{L,R\}, \quad (4)$$

where P is defined in Eq. (1) and N denotes the neighborhood size.

The value of V implies the degree of $L(i,j)$ satisfying the disparity d and the range of ψ is $[0,1]$. With V and ψ, we format the proposed skin
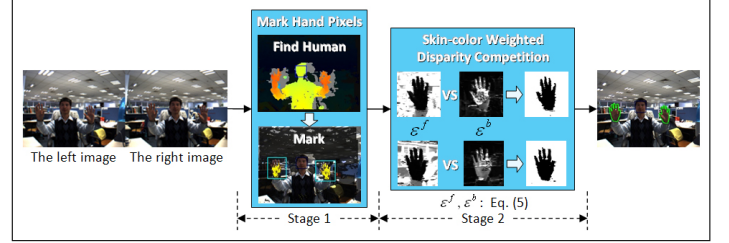

Figure 1: The framework of the proposed method.

color weighted disparity competition for the pixel $L(i,j)$ as

$$\begin{cases} \tilde{H}(i,j) = \begin{cases} 0, & if \ \varepsilon^h < \varepsilon^b \ and \ E(i,j) == 0 \\ 1, & otherwise \end{cases} \\ where \\ \varepsilon^h(i,j) = \min_{d \in \chi^h} \exp^{\frac{\left(1 - \min(\psi_L(i,j,N^h), \psi_R(i,j+d,N^h))\right) - (1-T_s)}{0.5 \times (1-T_s)}} \times \left(V\left(L(i,j),d,N^h\right) + \alpha\right) \\ \varepsilon^b(i,j) = \min_{d \in \chi^b(i,j)} \exp^{\frac{\psi_L(i,j,N^b) - T_s}{0.5 \times T_s}} \times \left(V\left(L(i,j),d,N^b\right) + \alpha\right) \\ E(i,j) = \begin{cases} 0, & if \ \psi_L(i,j,N^h) \geq T_s \ and \ \varepsilon^h \leq T^h \\ 1, & otherwise \end{cases} \end{cases}, \quad (5)$$

where

1. The verification error $\varepsilon^h(i,j)$ or $\varepsilon^b(i,j)$ is employed for verifying the degree that $L(i,j)$ satisfies the disparity limit of the hand or the background respectively:

   - Less value means more satisfying,
   - The neighborhood size of N^h or N^b is typically set to 2, T_s is a skin probability threshold and α is a small tuning constant;

2. $E(i,j)$ is employed for excluding the non-hand pixels ($E(i,j) = 1$) and consequently contributes to the low false acceptance rate;

3. $\tilde{H}(i,j)$ is employed to validate a hand pixel by competition:

   - $\tilde{H}(i,j) = 0$ means that $L(i,j)$ is a hand pixel since it satisfies the disparity limit of the hand more than of the background.

The exponential term in $\varepsilon^h(i,j)\backslash\varepsilon^b(i,j)$ results from the skin color probability, and is used as the weight term to affirm the hand\background pixels as well as reversely relieve most of the background\hand pixels in $\varepsilon^h\backslash\varepsilon^b$. The weight strategy improves greatly the robustness of the disparity competition $\tilde{H}(i,j)$ validating a hand pixel.

Finally, the proposed method segments the hand in each ROI by performing the morphological operation of opening following by searching the maximal connected region in the related ROI.

Note that only the step of the skin color weighted disparity competition (Eq. (5)) is necessary for the successive frames when ROIs tracking is incorporated, since all ROIs, the skin color probability and the hand disparity limit can be inferred or transferred from the previous frame.

Our experiments demonstrates the efficiency of the proposed method. The average segmentation accuracy of the proposed method is 83.8% and the frame rate is 19.2 fps on a Lenovo ThinkPad T400 with 2.5 GHz Dual Intel CPU and 2GB RAM.