Visual Tracking via Locality Sensitive Histograms

Yufeng Jiang

1. Illumination Invariant Features

In practice, most of the applications use normalized histograms. A normalization step, including a summation operation over all bins, for obtaining the normalization factor is required at each pixel location. A brute-force implementation of this summation operation is obviously $O(B)$ per pixel. Nevertheless, let n_p denote the normalization factor at pixel p shown at [1]:

$$n_p = \sum_{b=1}^{B} \mathbf{H}_p^E(b) = \sum_{q=1}^{W} \alpha^{|p-q|} \cdot (\sum_{b=1}^{B} Q(\mathbf{I}_q, b))$$

$$= \sum_{q=1}^{W} \alpha^{|p-q|} \tag{1}$$

Hence, the computation of the normalization factor n_p is independent of the number of bins B. In addition, n_p can also be computed in the same way:

$$n_p = n_p^{left} + n_p^{right} - 1, \tag{2}$$

$$n_p^{left} = 1 + \alpha \cdot n_{p-1}^{left}, \tag{3}$$

$$n_p^{right} = 1 + \alpha \cdot n_{p+1}^{right}. \tag{4}$$

In this case, the normalization factors can be computed in $O(1)$ per pixel. Furthermore, the normalization factors are indenpendent of the image content and thus can be precomputed to further reduce the computational complexity.

The algorithm presented in Eqs. 2-4 used at [1] is developed for 1D images. However, its extension to multidimensional images is straightforward. Authors simply perform the proposed 1D algorithm separately and sequentially in each dimension. Fig. 1 cited from [1] shows the speed of the proposed algorithm for computing the locality sensitive histogram for a 1 MB 2D image with respect to the logarithm of the number of bins B. Note that it is as efficient as the integral histogram.

Images taken under different illumination conditions have drastic effects on object appearance. Under the assumption of affine illumination changes, they can synthesize images of the scene presented in Fig. 2 shown at [1].
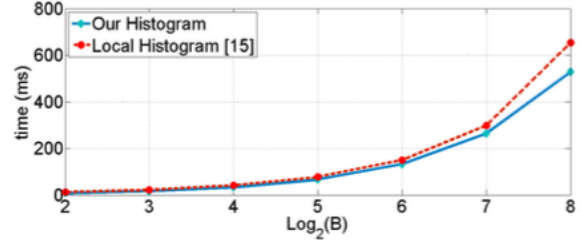

Figure 1. Speed comparison. The red dash curve shows the runtime of computing the local histograms from the integral histogram for a 1 MB image w.r.t. the logarithm of the number of bins B. The blue solid curve shows the runtime of computing the proposed locality sensitive histograms. Note that the speed of the two algorithms are very close and can be computed in real time when the number of bins is relatively low. For instance, when $B = 16$, the locality sensitive histograms can be computed at 30 FPS.
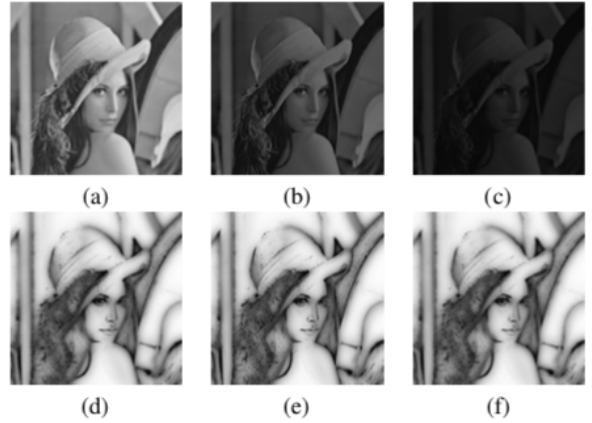

Figure 2. Illumination invariant features. (a)-(c) are input images with different illuminations. (d)-(f) are the corre- sponding outputs.

As the corresponding pixels have different brightness values, the intensity of the image cannot be considered as invariants for matching.

2. Tracking via Locality Sensitive Histograms

Based on the location of the target object center in the previous frame, authors aim to locate the new center location within the search region. Exhaustive search within the search region is performed in this work, where every pixel is considered as a candidate target center. Each region at the candidate location is matched correspondingly against the template to produce a score, based on the Earth Mover's Distance (EMD) [2].

Although the conventional EMD is computational expensive, the EMD between two normalized 1D histograms can be computed in time linear in the number of bins [3]. It is equal to the i_1-distance between their cumulative histograms [4]. The proposed locality sensitive histogram is normalized. The distance between histograms can be computed as Eq. 5 used in [1].

$$d(S_1, S_2) = \sum_{b=1}^{B} |C_1(b) - C_2(b)|, \qquad (5)$$

Where S_1 and S_2 are the corresponding regions of the template (centered at $p1$) and the candidate (centered at $p2$) in the current frame. C_1 and C_2 are the cumulative histograms of $\mathbf{H}_{p1}^{E}$ and $\mathbf{H}_{p2}^{E}$. They are defined as: $C(b) = \sum_{1}^{b} \mathbf{H}_{p}^{E}(b)$.

References

[1] S. He, Q. Yang, R. W. H. Lau, and J. Wang. Visual tracking via locality sensitive histograms. In *CVPR*, 2013. 1, 2

[2] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. *IJCV*, 40(2):99–121, 2000. 2

[3] S. Shirdhonkar and D. Jacobs. Approximate earth mover's distance in linear time. In *CVPR*, 2008. 2

[4] M. Werman, S. Peleg, and A. Rosenfeld. A distance metric for multidimensional histograms. *CVGIP*, 32(3):328–336, 1985. 2