are $m = m_L + m_{uv}$. We determine the color histograms in a rectangular region and describe the color histograms as $\mathbf{p} = \{p^{(u)}\}_{u=1,\ldots,m}$.

To calculate a similarity between candidate color histograms and the reference histogram, we therefore need a similarity measure on color histograms. A commonly used measure is based on the Bhattacharyya coefficient [7]. The distance between two m bins normalized histogram $\mathbf{p} = \{p^{(u)}\}_{u=1,\ldots,m}$ and $\mathbf{q} = \{q^{(u)}\}_{u=1,\ldots,m}$ is defined as

$$d[\mathbf{p},\mathbf{q}] = \sqrt{1 - \sum_{u=1}^{m} \sqrt{p^{(u)} q^{(u)}}} \quad . \tag{1}$$

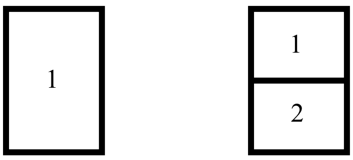
When two normalized histograms are a perfect match, we obtain $d = 0$.

Generally, a region to calculate histogram is represented as Figure 1 (a). The histogram calculated by this region has no spatial information. This can be a problem in tracking a color object because a spatial layout of target is beneficial. Therefore, we use a multi-part color histogram [5], [6]. This is a color histogram divided into sub-regions. We choose a multi-part histogram divided into two sub-regions as Figure 1 (b), because multi-part histogram has advantage of spatial color layout such as a human (e.g. one's colors of the shirt and trousers).

Hence the distance between two multi-part histograms $\mathbf{p} = \{\mathbf{p}_1, \mathbf{p}_2\}$, $\mathbf{q} = \{\mathbf{q}_1, \mathbf{q}_2\}$ is calculated using the average of the Bhattacharyya distance (in equation (1)) as

$$d_{multi}[\mathbf{p},\mathbf{q}] = \frac{\sum_{i=1}^{2} d[\mathbf{p}_i, \mathbf{q}_i]}{2} \tag{2}$$

where $\mathbf{p}_i$ and $\mathbf{q}_i$ are i-th sub-region histogram shown as Figure.1 (b) respectively, and $d[\mathbf{p}_i, \mathbf{q}_i]$ is calculated by equation (1).



(a) (b)

Figure 1. Histogram representations.

2.2. Combine adaptive target color representation and particle filtering

To track an object, we estimate an object's state $\mathbf{s}_t$ at time t. We define the object's state at time t as

$$\mathbf{s}_t = \{x_t, y_t, scale_t\} \tag{3}$$

where x_t, y_t specify the center of the rectangular region and $scale_t$ is the size of the rectangular region at time t. We use particle filtering to estimate the object's state.

Particle filtering (or Condensation) [1] is a time series filter for estimating a state. It was originally developed to track objects in clutter. The idea of particle filtering is to approximate the probability distributions of a state by a weighted sample set. The sample set at time t can be written as $\{(\mathbf{s}_t^{(i)} \pi_t^{(i)}), i = 1 \ldots N\}$, where $\mathbf{s}_t$ is the object's state defined above, π_t is its weight, and N represents the number of samples. The sample set is propagated by a system model. Then, the new generated sample set is weighted in terms of the observations, and the mean state $E[\mathbf{s}_t]$ is estimated as

$$E[\mathbf{s}_t] = \sum_{i=1}^{N} \pi_t^{(i)} \mathbf{s}_t^{(i)} \quad . \tag{4}$$

Thus, particle filtering generates a new sample set at each step, we obtain estimated an object's state.

We apply particle filtering to our tracking method. The key elements of a tracker using particle filtering are propagation using a system model and observation of a sample set.

We choose random walk as a system model. The reason we use this system model is that we assume various situations in case of occurring occlusion, clutters, or target appearance change by a moving camera. The situation such as an image form a moving camera does not apply a first order model [4], [5] to a system model. Random walk is represented as

$$\mathbf{s}_t = \mathbf{s}_{t-1} + \mathbf{w}_{t-1} \tag{5}$$

where $\mathbf{w}_{t-1}$ is a system noise. We use the uniform distribution which has zero mean as a system noise $\mathbf{w}_{t-1}$. We define that the range of the uniform distribution $\mathbf{w}_{t-1}$ is changed by the rectangular region size covering the object. When the rectangular region size is large, the velocity of the rectangular region on image is high, otherwise it is low. Therefore, the range of the uniform distribution $\mathbf{w}_t$ is described as below.

$$\begin{pmatrix} w_t^x \\ w_t^y \\ w_t^{scale} \end{pmatrix} = \begin{pmatrix} \frac{1}{a} width_{t-1} \\ \frac{1}{a} width_{t-1} \\ \frac{1}{b} width_{t-1} \cdot height_{t-1} \end{pmatrix} \tag{6}$$

Where, w_t^x and w_t^y are the range of the center x, y of the rectangular region respectively, w_t^{scale} is the range of the rectangular size at time t, $width_{t-1}$ and $height_{t-1}$ are the width, height of the rectangular at time $t-1$ respectively, and a, b are parameters. Thus, to predict the state of the sample set, each sample is propagated by equation (5).

To weight the sample set, we use color histograms as observation. Observation accuracy depends on tracking object's features. As we describe multi-part histogram above, it is more robust than single-part histogram. However, multi-part histogram has a disadvantage which is less robust than single-part histogram when object's color is plain. This disadvantage is shown in Section 3 in detail. Hence, we use multi-part color histogram and single-part color histogram adaptively when a user sets the object in initializing step. Determination to use which multi-part color histogram or single-part color histogram is described as

$$d[\mathbf{p}_1, \mathbf{p}_2] > T \tag{7}$$

where $\mathbf{p}_1$ and $\mathbf{p}_2$ are sub-region histograms of multi-part histogram represented by Figure 1 (b), $d[\cdot,\cdot]$ is the Bhattacharyya coefficient in equation (1), and T is a threshold of determination. According to (7), our method could track an object having the relative position of colors or not (See in Section 3). To weight each sample of the set, we calculate the Bhattacharyya coefficient d . We assume that the sample set's weight distribution is Gaussian, and as color distributions of samples is similar to the reference color distribution (d is small), the samples have large weights. Therefore, each sample weight is calculated as Gaussian with variance σ by equation below.

$$\pi_t^{(i)} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{d^2}{2\sigma^2} \right) \quad (8)$$

Next, to obtain the object's state, we calculate the mean state $E[\mathbf{s}_t]$ of the sample set by equation (4).

Finally, we select the new sample set from the sample set. This is called re-sampling. Re-sampling ensure the accuracy of the probability distribution of the state. Its technique is that the higher the weight of a sample, the more likely it is to be duplicated, in contrast the lower the weight, the more likely it is to be eliminated. We use re-sampling algorithm described in [1]. See [1] for details. We show our tracking procedure in Figure 2.

Initializing

(1) select an object to track by rectangular region

(2) calculate sub-regions (in Figure 1 (b))color histograms $\mathbf{p}_1$ and $\mathbf{p}_2$ in the rectangular region

(3) detect to use which multi-part color histogram or single-part color histogram by equation (7) if this equation is fulfilled, the flag is equal to 1

(4) calculate the reference histogram $\mathbf{p}$

if the flag is equal to 1, calculate multi-part color histogram $\mathbf{p} = \{ \mathbf{p}_1, \mathbf{p}_2 \} = \{ p_1^{(u)}, p_2^{(u)} \}_{u=1,\dots,m}$

else, calculate single-part color histogram $\mathbf{p} = \{ p^{(u)} \}_{u=1,\dots,m}$

Tracking

Given the sample set $\{(\mathbf{s}_{t-1}^{(i)}, \pi_{t-1}^{(i)}), i = 1 \dots N\}$ and the reference histogram $\mathbf{p}$. Color distributions of each sample and their Bhattacharyya coefficient d are obeyed as follows,

If the flag is equal to 1, they are calculated as multi-part color histogram, and d is calculated by equation (2),

else, they are calculated as single-part color histogram and d is calculated by equation (1).

(1) propagate each sample $\mathbf{s}_t^{(i)}$ from the sample set $\mathbf{s}_{t-1}^{(i)}$ by equation (5)

(2) calculate each sample weight $\pi_t^{(i)}$

(a) calculate color distributions of each sample $\mathbf{s}_t^{(i)}$

(b) calculate the Bhattacharyya coefficient d for each sample $\mathbf{s}_t^{(i)}$

(c) weight each sample $\mathbf{s}_t^{(i)}$ by equation (8)

(3) estimate the mean object state by equation (4)

(4) select next new N samples $\{(\mathbf{s}_{t+1}^{(i)}, \pi_{t+1}^{(i)}), i = 1 \dots N\}$ from the sample set $\{(\mathbf{s}_t^{(i)}, \pi_t^{(i)}), i = 1 \dots N\}$

(5) repeat (1) ~ (4) during tracking object

Figure 2. Object tracking procedure.

3. Experiments

In this section, we evaluate our proposed method. We first show need of the determination to use which multi-part histogram or single-part histogram on video sequences in outdoor scene. Then, we present the experimental results of our proposed method on these video sequences in outdoor scene.

In experiments, we use eight video sequences under different conditions in outdoor scene. In Sequence 1-4, a walking person is captured by a moving camera. These Sequences include target appearance changes and background changes rapidly. In Sequence 5-8 walking persons are captured by a fixed camera. These Sequences include background clutters and target's occlusion. Furthermore, in Sequence 1, 2, 5, and 6, the target has the relative position of colors of the shirt and trousers. In Sequences 3, 4, 7, and 8, the target has plain color. The image size of these video sequences is 640×480. The experiments are implemented using a Windows PC with Intel Pentium4 3.0 GHz and 1.0 GB memory RAM.

The parameters are the same for all test video sequences. They are described as follows. The bins of L histogram are $m_L = 10$, the bins of uv histogram are $m_{uv} = 10 \times 10$. Total bins are $m = m_L + m_{uv} = 110$. The numbers of the sample set are $N = 200$. System noise parameters in equation (6) are respectively $a = 3$, $b = 10$. The variance to weight each sample in equation (8) is $\sigma = 0.1$. The threshold of determination to use multi-part histogram in equation (7) is $T = 0.4$.

The results are shown in Table 1. Table 1 shows the performance of tracking an object using single-part color histogram, multi-part color histogram, and our proposed method (the adaptive target color representations based on multi-part or single part color histogram) with particle filtering on eight video sequences.

First, we exhibit need of the determination to use which multi-part histogram or single-part histogram shown in Section 2.2. According to Table 1 using single-part histogram and multi-part histogram, we found that an advantage and a disadvantage of multi-part histogram. The advantage is that multi-part histogram is more robust than single-part histogram in video sequences except for Sequence 3, 4, 7, and 8 (object' color is plain). These results show that multi-part histogram is more robust than single-part histogram because of using spatial information. In the other hand, the disadvantage is that the performance using multi-part histogram is less than that of using single-part histogram when the object's color is plain (such as Sequence 3, 4, 7, and 8). Therefore, a tracker using multi-part histogram is inferior to one using single-part histogram in tracking a plain color object. Furthermore, in experiments, the average computing time of multi-part histogram takes 10 milliseconds more than that of single-part histogram. This shows that single-part histogram is beneficial when object's color is plain, because one's computing time is no more than that of multipart histogram even though one's performance is equal to that of multi-part histogram. Thus, we propose the adaptive target color representation based on multi-part histogram and single-part histogram.

Then, we evaluate the performance of our proposed tracking method. According to Table 1, our proposed method can track an object which has the relative position of colors (e.g. the shirt and trousers) or plain color by selecting multi-part or single-part color histogram adaptively. Thus, our proposed method can solve the problems of multi-part histogram and single-part histogram. The tracking results using our proposed method are showed in Figure 3 and 4. Figure 3 and 4 shows tracking results of Sequence 2 and 6 respectively. In Sequence 2, the person's appearance changes by a moving camera. In Sequence 6, the person is walking in case of occlusion and clutters. Our tracking method successfully tracks the object in case of these severe situations.

Table 1. Comparison of tracking accuracy results for different target representations.

Sequence		Numbers of frames of tracking successfully		
No	Total frames	Single-part	Multi-part	proposed method
1	788	736	788	788
2	534	346	534	534
3	1315	1315	1315	1315
4	1308	1308	815	1308
5	300	171	300	300
6	269	158	269	269
7	377	337	336	337
8	229	229	229	229

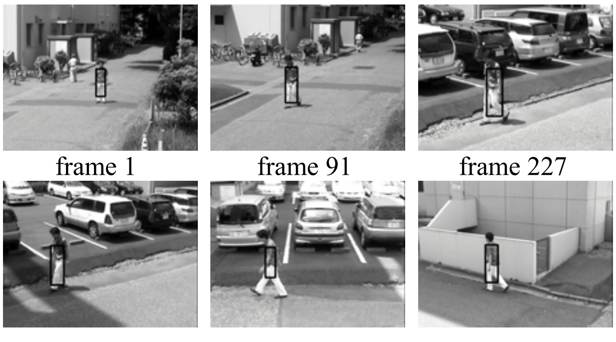


Figure 3. Object tracking result of Sequence 2.

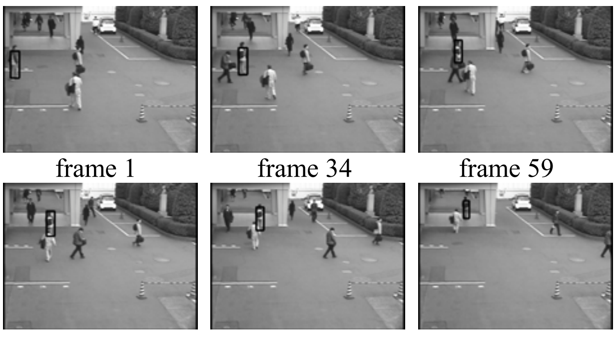


Figure 4. Object tracking result of Sequence 6.

4. Conclusion

We propose the object tracking method based on the adaptive target color representation based on multi-part color histogram and single-part color histogram and particle filtering. We show the advantage of robustness and the disadvantage of multi-part histogram on video sequences in outdoor scene. The advantage is that it is more robust than single-part histogram using its spatial information. The disadvantage is that it is less robust than single-part histogram when object's color is plain because of no spatial information. Therefore, we propose the tracking method using multi-part color histogram and single-part color histogram. In initializing an object to track, our tracking method determinates to use which multi-part histogram or single-part histogram. Furthermore, by combining the adaptive color representation and particle filtering, our proposed tracker could track an object in cases of clutters or occlusion. In experiments, our tracking method successfully tracks the object having multi-color (e.g. relative position of the shirt and trousers) or plain color on video sequences in cases of target's appearance change or occlusion.

Future work includes using an object's other features (e.g. edge). We are planning to combine color information and other features to solve the problem that a color distributions based tracker is poor if a similar looking object is nearby.

Acknowledgments

We would like to thank Koito Industries, Ltd for helpful suggestions and assistance with experiments.

References

[1] M. Isard and A. Blake: "Condensation- conditional density propagation for visual tracking", *Int. J. Computer Vision*, vol.29, no.1, pp.5-28, 1998.

[2] G.R.Bradski: "Computer vision face tracking as a component of a perceptual user interface," *In Workshop on Applications of Computer Vision*, pp.214-219, 1998.

[3] D. Comaniciu, V. Ramesh and P. Meer: "Real-Time Tracking of Non-Rigid Objects using Mean Shift," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.142-149, 2000.

[4] K. Nummiaro, E. K. Meier and L. V. Gool: "An Adaptive Color-Based Particle Filter," *Image and Vision Computing*, vol.21, no.1, pp.99-110, 2002.

[5] P. Perez, C. Hue, J. Vermaak, and M. Gangnet: "Color-Based Probabilistic Tracking," *in Proceeding of the European Conference on Computer Vision*, pp.661-675, 2002.

[6] E. Maggio and A. Cacallaro: "Multi-part target representation for color tracking," *in Proceeding of IEEE International Conference on Image Processing*, vol.1, pp.729-732, 2005.

[7] F. Aherne, N. Thacker, and P. Rockett: "The Bhattacharyya metric as an absolute similarity measure for frequency coded data," *Kybernetica*, vol.32, no.4, pp.1-7, 1997.