IPSJ Transactions on Computer Vision and Applications Vol. 2 145–155 (Nov. 2010)

Research Paper

Efficient Acquisition of Human Existence Priors from Motion Trajectories

Hitoshi Habe,[†1] Hidehito Nakagawa[†1,*1] and Masatsugu Kidode[†1]

This paper proposes a method for acquiring the prior probability of human existence by using past human trajectories and the color of an image. The priors play an important role in human detection as well as in scene understanding. The proposed method is based on the assumption that a person can exist again in an area where he/she existed in the past. In order to acquire the priors efficiently, a high prior probability is assigned to an area having the same color as past human trajectories. We use a particle filter for representing and updating the prior probability. Therefore, we can represent a complex prior probability using only a few parameters. Through experiments, we confirmed that our proposed method can acquire the prior probability efficiently and use it to realize highly accurate human detection.

1. Introduction

In recent years, human beings have been increasingly subjected to visual surveillance. Because manned observation is unfeasible, sophisticated techniques such as Ref. 1) that can extract important and useful information automatically are required. In particular, understanding human activities is one of the most essential and important issues in visual surveillance.

In order to understand human activities from videos, many researchers have been considering the prior probability of human existence in the context of an observed scene. The priors have several applications, as given below.

First, they can be used to improve the performance of human detectors and human trackers. Occasionally, it is difficult to detect and track walking persons using only the appearance within a local image patch because even well-trained human detectors fail when there is no difference between the image patterns of persons and other objects in their environments. If the probability of human existence at each position in an image is available, it is possible to avoid over- and miss-detection and improve the performance of human detectors and trackers [2)–4)].

Additionally, the distribution of the priors reveals considerable information about an observed scene. For example, if there are some locations that attract people, the spatial distribution would be uneven. Similarly, temporal variations may indicate that the flow of walking persons changes for some reason. Such information is useful for providing adequate services such as guidance for visitors.

Some research groups have already proposed methods for obtaining the priors of human existence; these methods can be divided into several categories. First, if the map of a scene is known, it can be used to derive priors directly [2)]. Some methods that estimate the geometric structure of a scene [5)] can be used for accurate human detection [3)]. Because the geometric structure of a scene directly affects human actions in a scene, these methods are quite natural and straightforward. If no information about the geometric structure is available, we can acquire the priors from observed human trajectories. As described in Ref. 4), by accumulating trajectories in long sequences, it is possible to estimate the priors in a scene.

In this study, we propose an efficient method that acquires the priors of human existence from time-series images of a scene. This method employs the human trajectories and color information of the images. If a few static cameras are used for surveillance and the structure of observing scenes remains unchanged, it is easy to obtain the priors for each camera manually. However, it is not feasible to obtain the priors for thousands of cameras or to maintain changing priors of the scenes when the camera changes its viewing direction or the scene structure varies. The proposed method is intended to be applied to such situations.

As described above, human trajectories are a cue that can be used to estimate the prior particularity when no geometric structure is available. However, a large number of trajectories are required for accurate estimation. For example, if people walk along wide roads such as those shown in **Fig. 1** and **Fig. 2**, the motion trajectories will exhibit a sparse distribution on the road. Hence, in

†1 Nara Institute of Science and Technology
*1 Presently with KEYENCE Corporation

Fig. 1 Scene 1.

Fig. 2 Scene 2.

order to obtain the optimal priors, which should be uniform on the road, we have to collect a large number of trajectories. Therefore, we also employ the color information of the images. We assume that pixels corresponding to the same region such as a road should have similar color. Higher priors will be assigned to similarly colored areas having past motion trajectories.

Additionally, we use a particle filter for representing and updating the priors. This makes it possible to represent the complicated distribution of the priors and to adapt the distribution to scene changes such as the movement of background objects. Furthermore, it can capture the "dynamics" of the priors that would reflect the context of a scene, as described above.

2. Prior Probability of Human Existence

Before describing the proposed method, we introduce the definition of the prior probability distribution of human existence and describe how it can be used in practical applications.

2.1 Definition and Representation of Human Existence Priors

Our main objective is to understand large-scale events occurring in the real world in a manner similar to humans. To do so, we have to employ and integrate various types of information and knowledge efficiently. Among the available information, the "context" of a scene would play an important role, and some works that make use of the context have already been proposed. The literature mentioned in Section 1 are typical examples of such works. The context would have various meanings depending on the application and situation. The human existence prior is one of the fundamental features that describes the context of a scene.

The main factors that determine the priors are as follows:

Geometric Structure: People are naturally more likely to walk on horizontal planes in a scene, such as roads, than in other areas, such as walls or the roofs of buildings. Knowing such structures, i.e., *geometric structures*, in advance, helps in the acquisition of priors.

Semantic Structure: Some areas in a scene have special meaning; for example, a large number of people tend to move in and out of a structure near the entrance and exit. If there is an information board on a road, people are likely to walk near it. Structures that give meaning to a certain area are called *semantic structures*. In addition to the geometric structure shown above, such semantic knowledge provides valuable and meaningful information that can be used for obtaining priors.

While geometric structures have already been used for obtaining priors [2),3)], as discussed in Section 1, few studies have investigated the use of semantic structures. This might be because a variety of semantic structures are available and there is no definite method for obtaining and representing them.

In this study, as discussed in Section 1, we make use of human motion trajectories and image colors to obtain the human existence priors. We regard the color information as a fundamental feature that indicates the geometric structure. This is based on the assumption that a geometrically uniform region has a uniform color. On the other hand, the human motion trajectories can be regarded as a cue that can be used for estimating the priors derived from the geometric structure as well as the semantic structure, such as the flow of walking people.

As discussed in Section 1, although the geometric structure is an important cue for robust pedestrian detection, it is not always easy to obtain the geometric structure accurately, especially for outdoor scenes. In contrast, the human motion trajectories reflect both the geometric and the semantic structure and can be used for estimating the human existence priors. However, it is sometimes inefficient to estimate the priors solely by the past trajectories, as discussed in Section 1.

From this viewpoint, in this study, the proposed method estimates the priors by using both the image colors and the human motion trajectories.

Clearly, the priors $P(p)$ differ according to the position in an image. Therefore, we have to maintain the value of $P(p)$ at each position. However, maintaining each value of $P(p)$ is inefficient, and these values have redundancy in a spatial domain. Therefore, we use the framework of a particle filter, that is, the prior distribution is approximated by the density of particles. This enables us to update the distribution efficiently.

2.2 Application of Human Existence Priors

As already discussed in Section 1, the priors $P(p)$ can be used for improving the performance of human detectors and human trackers. Human detectors often make use of an intensity pattern in a local image window [6]. In other words, they do not consider information of other areas, such as the co-ocurrence with other objects or the 2D or 3D positions in a scene. The use of human existence priors $P(p)$ will supplement the use of human detectors, and it is expected to lead to an improvement in their performance, as demonstrated in some studies [2)–4)].

In the framework of Bayes' rule, this can be written as follows:

$$P(p|Y) = \frac{P(Y|p)P(p)}{P(Y)}, \tag{1}$$

where Y denotes an observed image and p indicates the existence of a person at a certain position. Obviously, both the prior $P(p)$ and the likelihood $P(Y|p)$, which can be estimated by the human detectors, provide the posterior probability $P(p|Y)$ that is used for determining whether a person exists or not. Bayes' rule also indicates the importance of the human existence prior $P(p)$.

Additionally, the distribution of the priors reveals considerable information about an observed scene. For example, suppose a system provides adequate information according to the condition of a walking person and the situation of a scene. Toward this end, we have to extract comprehensive features that characterize the condition and situation from the observed trajectories. We believe that the temporal variation of the prior distributions would be one such feature. In this study, although we do not show applications and results supporting this claim, we believe that our method is applicable to human detection as well as various other applications. For example, the temporal variation would show the time-varying attention of pedestrians walking on the road. This would be applicable when considering marketing strategies.

3. Efficient Acquisition of Human Existence Priors

We describe our proposed method in this section.

3.1 Overview of Proposed Method

Let $Y_t = \{y_1, y_2, \cdots, y_t\}$ denote a sequence of observed images from time 1 to t and Ψ_t be a group of detected human trajectories in Y_t. The proposed method estimates the priors using them. In order to make it clear what is used for the estimation, we denote the priors $P(p_t|Y_{t'}, \Psi_{t'})$ that represent the priors at time t estimated by using observed images and human trajectories from time 1 to t'.

Figure 3 shows the process flow of the proposed method at a certain time t. Before starting the processing at time t, the priors $P(p_{t-1}|Y_{t-1}, \Psi_{t-1})$ were obtained. Using this method, we first perform human detection for the observed image y_t at time t ((1) in Fig. 3). As shown in Eq. (1), the detection is carried out by applying a certain threshold to the posterior probability $P(p|Y)$ obtained by both the priors and a human detector. As the result of the human detection, we obtain the human trajectories Ψ_t. Then, the priors $P(p_{t-1}|Y_{t-1}, \Psi_{t-1})$ are updated to $P(p_t|Y_t, \Psi_t)$ using Y_t and Ψ_t ((2) in Fig. 3).

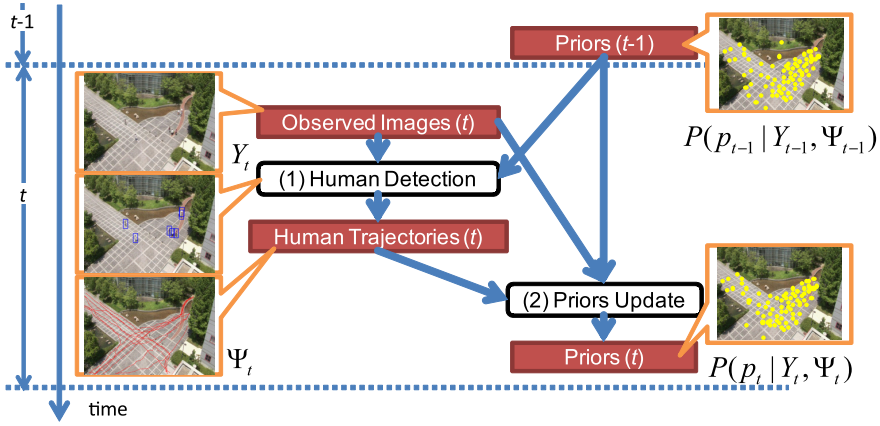


Fig. 3 Efficient acquisition of human existence priors—overview of proposed method.

These processes are iteratively conducted so that the priors adapt to variations in a scene. Note that our current implementation uses uniform priors at initial time $t = 1$.

Updating the priors is a crucial step in the proposed method. As mentioned above, we employ the framework of a particle filter in order to represent the complicated distribution of the priors efficiently and to adapt to the temporal variation of the priors' distributions efficiently. This is described in detail in the following section.

3.2 Representing and Updating Priors Using Particle Filters

Figure 4 shows how the priors are updated using the particle filter. As mentioned in the previous section, the prior $P(p_{t-1}|Y_{t-1}, \Psi_{t-1})$ is updated by using observed images Y_t and detected human motion trajectories Ψ_t. We represent the priors by the spatial distribution of a set of weighted samples $S_{t|t'} = \{s^i_{t|t'}|i = 1, 2, \ldots, N\}$ where $s^i_{t|t'}$ denotes the i-th particle at time t estimated from the data until time t'. Namely, $P(p_t|Y_t, \Psi_t)$ is represented by the samples $S_{t|t}$. In keeping with the normal usage of the particle filter, the update is conducted as follows:

(1) Estimating Current State

From the previous sample set $S_{t-1|t-1}$, we estimate the current set $S_{t|t-1}$. This estimation is performed using a state transition model $S_{t|t-1} = F(S_{t-1|t-1}, Y_{t-1}, \Psi_{t-1})$. The estimated sample set denotes the priors $P(p_t|Y_{t-1}, \Psi_{t-1})$ that are derived from images and trajectories until time $t-1$.

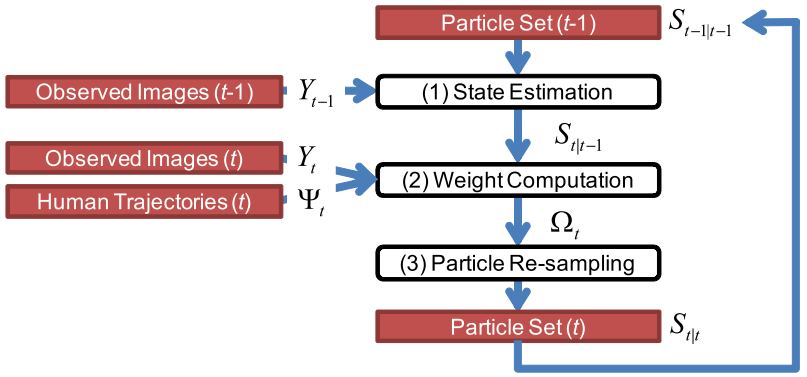


Fig. 4 Updating priors using particle filters.

(2) Computing Weight of Each Particle

Then, we compute a set of weights $\Omega_t = \{\omega^i_t\}$ for the estimated current particles $S_{t|t-1}$ using the observation at time t. We introduce a weight function $\omega^i_t = H(s^i_{t|t-1}, Y_t, \Psi_t)$ for the computation.

(3) Re-sampling According to Ratios of Weights

Finally, we derive a particle set $S_{t|t}$ by re-sampling $S_{t|t-1}$ according to the weights ω^i_t. Specifically, the number of new particles located at the same position as the particle $s^i_{t|t-1}$ is determined so that it is proportional to the ratios of the weights $\frac{\omega^i_t}{\sum_i \omega^i_t}$.

The obtained $S_{t|t}$ represents the probability distribution $P(p_t|Y_t, \Psi_t)$, which will be used as a prior at the next time $t+1$.

In the above procedures, the state transition model $S_{t|t-1} = F(S_{t-1|t-1}, Y_{t-1}, \Psi_{t-1})$ makes use of the color information observed in an image. The weights ω^i_t are determined mainly depending on past human motion trajectories. These are described in detail in the following sections.

3.2.1 Estimating Current State Using Color Information

As described in the previous sections, we assume that people are likely to appear in the regions that have a color similar to that of the regions they have already passed through. Based on this assumption, we move the i-th particle at time $t-1$, which has coordinates $\phi^i_{t-1} = (x^i_{t-1}, y^i_{t-1})$ as its state, to a position having a color similar to that of the current position. **Figure 5** illustrates this process.

Let c denote a 3D vector that represents the RGB color, c^i_t be the color of the position where particle $s^i_{t|t}$ exists, and c^ϕ_t be the color at position ϕ.

First, we obtain color c^i_{t-1} corresponding to particle $s^i_{t-1|t-1}$. Before obtaining the current state, i.e., position, of the particle, we randomly select a color similar to c^i_{t-1} as

$$c'^i_t = c^i_{t-1} + k_t, \tag{2}$$

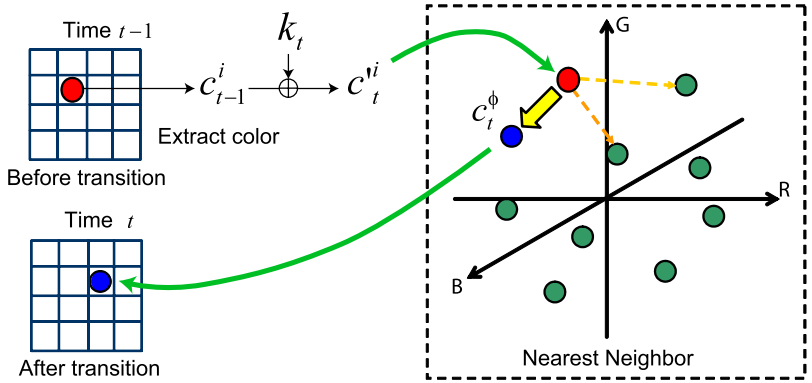


Fig. 5 Estimating current state using color information.

where k_t is a 3D vector in which each component is a small number such as $k_t \in [-5, 5]$. Then, we find a position ϕ^i_t where the color is the same as or similar to that of c'^i_t by minimizing the L2-norm

$$d^i = \left| c^\phi_t - c'^i_t \right|, \tag{3}$$

where c^ϕ_t denotes the colors of all pixels in the observed image. When we find $\phi^i_t = \arg\min_\phi d^i$ for a particle at time t, an approximate nearest neighbor (ANN) search [7] is applied for efficient computation. As a result of this processing, the position ϕ^i_t is selected as the current state, i.e., position, of the particle $s^i_{t|t-1}$.

3.2.2 Computing Weight of Each Particle Using Past Motion Trajectories

Next, we compute a weight for each particle. This weight shows how the particle is likely to exist; in other words, how people are likely to exist at a position that corresponds to the particle.

To compute the weight, we define a type of distance between a particle and past motion trajectories. The distance depends on both the Euclidean distance between the particle and the trajectories and the difference between their colors. This is described in detail in the following sections.

3.2.2.1 Euclidean Distance to Past Trajectories

As we introduced in Section 3.2, a set of trajectories observed between time 1 and time t is denoted by $\Psi_t = \{\psi_1, \psi_2, \cdots, \psi_{N^{trj}_t}\}$, where ψ_i is a respective trajectory and N^{trj}_t is the number of trajectories observed. Let ϕ^i_t be the position of the i-th particle at time t.

Because the weight will be large at a position where people are likely to exist, the weight $\mathcal{L}^i_{trj}(t)$ derived from the Euclidean distance is given as:

$$\mathcal{L}^i_{trj}(t) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{{d^i_{trj}(t)}^2}{2\sigma_1^2} \right), \tag{4}$$

where $d^i_{trj}(t)$ denotes the minimum distance between the i-th particle and the positions in the observed trajectories Ψ_t, defined as:

$$d^i_{trj}(t) = \min_{\psi_i \in \Psi_t,\, \chi_i \in \psi_i} \left| \phi^i_t - \chi_i \right|, \tag{5}$$

where $\chi_i \in \psi_i$ is a respective position included in the trajectory ψ_i.

3.2.2.2 Difference in Color to Past Trajectories

In addition to the Euclidean distance, we incorporate the similarity of color information into the weight. This is because if only the distance to past trajectories is considered, more trajectories passing the road overall would be required to obtain adequate priors.

First, we segment an input image using the method described in Ref. 8) as shown in **Fig. 6** (a). Although this method would not provide accurate segmentation, it is not necessary for us to obtain accurate segments because here, the purpose of segmentation is to obtain roughly uniform regions in an image.

Then, when the motion trajectories are observed, we integrate a segment that corresponds to the trajectories, into one segment as shown in Fig. 6 (1). Finally, we assign a uniform weight that has a high value when a particle lies in the integrated segment, shown with green color in the bottom row of Fig. 6, and a low weight when the particle lies outside the region. This weight function can be written as

$$d^i_{seg} = Z(\phi^i_t), \tag{6}$$

$$\mathcal{L}^i_{color}(t) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{{d^i_{seg}}^2}{2\sigma_2^2} \right), \tag{7}$$

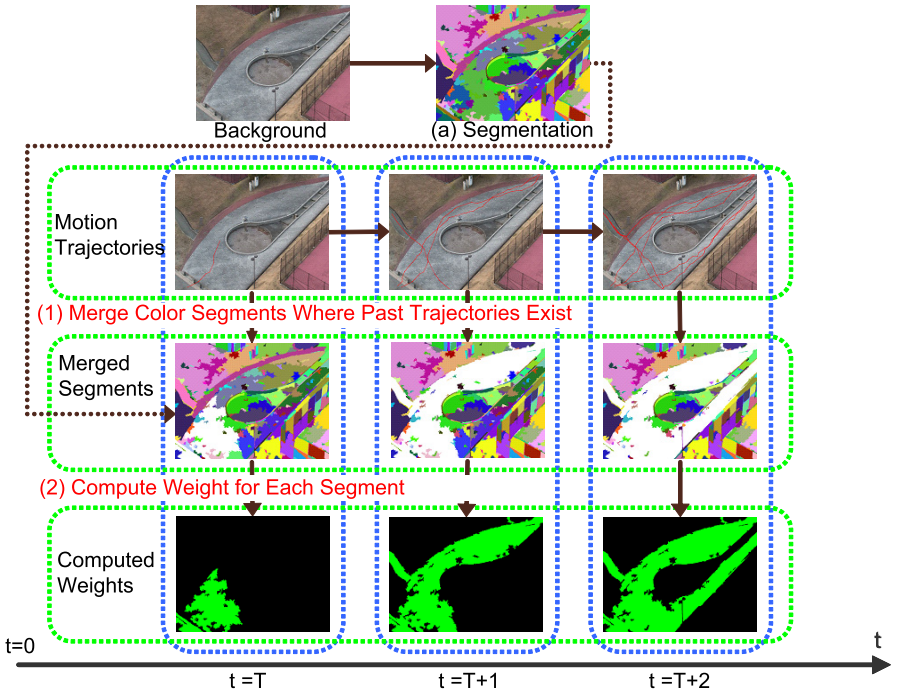


Fig. 6 Color region integration for computing weight of each particle.

where $Z(\cdot)$ denotes the difference derived from the color at the particle; if it lies in the integrated region, it will be small, otherwise it will be high. The function $Z(\cdot)$ is intended to able to represent the likelihood derived from the geometric structure. For example, if the height in a 3D space is available for each pixel, $Z(\cdot)$ would be the differences in the heights from past human trajectories. In the current implementation, however, it simply takes 0 for the integrated region and 1 otherwise.

Now we have two types of weight functions. Finally, these functions are integrated, as given by Eq. (8), and used for determining the weight of the i-th particle at time t:

$$\omega_t^i = \mathcal{L}_{trj}^i(t) + \mathcal{L}_{color}^i(t). \tag{8}$$

3.3 Computing Prior Using Particles

When we make use of the prior, a set of particles $S_{t|t-1}$ is transformed to the prior $P(p_t|Y_{t-1}, \Psi_{t-1})$ at position (x, y) as:

$$P(p_t|Y_{t-1}, \Psi_{t-1}) = \alpha K(x, y, \sigma_3) * \sum_i \rho(x, y, s_{t|t-1}^i), \tag{9}$$

where K denotes a Gaussian kernel that has variance of σ_3 and the operation $*$ denotes a convolution. ρ takes 1 if a particle $s_{t|t-1}^i$ exists at (x, y) and takes 0 otherwise. Their sum at a certain position gives the number of particle located there. Note that when we utilize the priors for certain applications, such as pedestrian detection shown in the next section, the relative difference of the priors among positions in an image plays an important role. Therefore, the scale factor α can be ignored.

4. Experiments

This section presents the experimental results that show the effectiveness of the proposed method.

As discussed in the previous section, we have to give parameters for the proposed method. Based on some preliminary experiments, we used $\sigma_1 = 30$, $\sigma_2 = 80$, $\sigma_3 = 150$, and the number of particles $N = 8000$ in the following experiments, unless otherwise stated.

4.1 Videos and Trajectories for Experiments

We conducted two kinds of experiments. In Experiment 1, we captured two outdoor videos for the experiments, as shown in Fig. 1 and Fig. 2. For each video, human motion trajectories were given manually, as shown in **Fig. 8** and **Fig. 10**. Using this data, we show the results of acquiring human existence priors and human detection using the acquired priors. This experiment aims to confirm the basic effectiveness of our method.

We then tested the proposed method under a more realistic scenario in Experiment 2. We made use of a longer video sequence, as shown in **Fig. 7**, which is part of the PETS 2006 benchmark data [9]. In this experiment, human motion trajectories were given by a simple pedestrian detector using the HOG features [6] and the SVM classifier [10].

Fig. 7 Scene 3 (PETS 2006 dataset).

4.2 Quantitative Comparison of Pedestrian Detection Results

For both Experiment 1 and 2, we quantitatively evaluated the effectiveness of the proposed method by comparing pedestrian detection results with a ground truth given manually.

Pedestrian detection is carried out by applying the detector described in the previous section. From Eq. (1), the posterior probability can be written as:

$$P(p|Y) \propto P(Y|p)P(p). \tag{10}$$

Here we assume that the likelihood $P(Y|p)$ is equivalent to the output of the SVM classifier used in the pedestrian detector. The priors $P(p)$ are given using three different methods, including the proposed method, as follows:

(1) Uniform Priors $P(p)$ take a constant value for an image. This is equivalent to performing pedestrian detection without considering priors.

(2) Priors from Trajectories Positions in past human trajectories are accumulated. Then, when performing human detection, the accumulated positions are examined and a high prior is assigned if the current position is on or near the past trajectories.

(3) Priors from Trajectories and Color — Proposed The priors are assigned using the method described in Section 3. The is the proposed method in this paper.

For quantitative comparison, we compute the precision, recall, and F-value, respectively defined as follows:

Fig. 8 Scene 1: human motion trajectories.

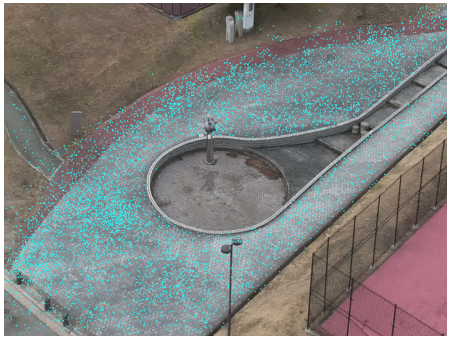


Fig. 9 Scene 1: particle distribution.

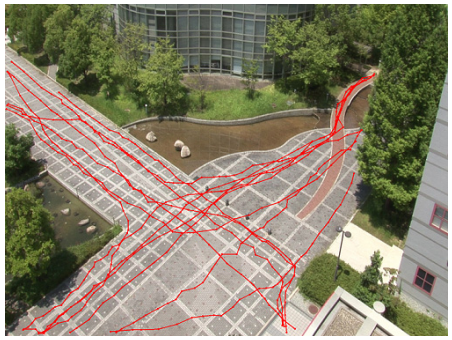


Fig. 10 Scene 2: human motion trajectories.

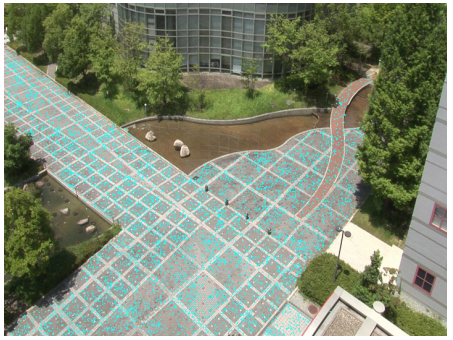


Fig. 11 Scene 2: particle distribution.

$$Precision = \frac{TP}{TP + FP}, \tag{11}$$

$$Recall = \frac{TP}{TP + FN}, \tag{12}$$

$$F = 2/\left(\frac{1}{Precision} + \frac{1}{Recall}\right), \tag{13}$$

where TP, FP, and FN denote True-Positive, False-Positive and False-Negative, respectively. It is evident from the definitions that larger values correspond to good performance. Note that in order to compute these values we need to apply a certain threshold to the product of $P(Y|p)$ and $P(p)$ in Eq. (10). The values

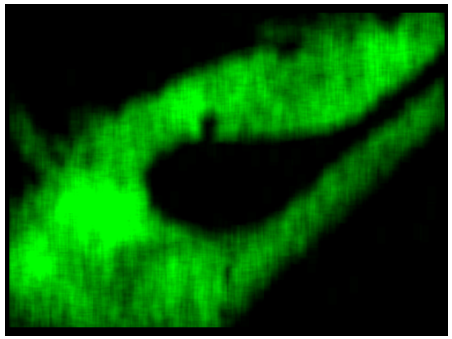


Fig. 12 Scene 1: human existence priors by proposed method.

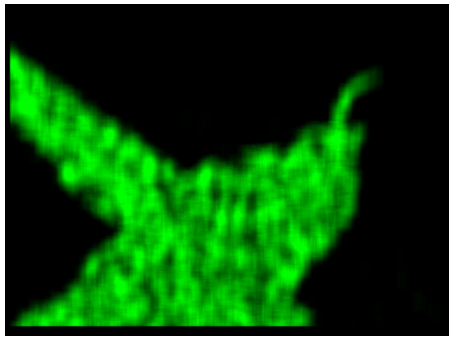


Fig. 13 Scene 2: human existence priors by proposed method.

shown in the following sections are the results when their F-values are the best among the results using various threshold values.

4.3 Experiment 1 — Estimation Results Using Manually Selected Human Trajectories

First, we show the results using manually selected human trajectories for 200 images. The distribution of the obtained particles are shown in **Fig. 9** and **Fig. 11**. From the results, we can see that particles are distributed not only on the past trajectories but also in the area that has a color similar to that of the trajectories. Note that there are 10 trajectories for Scene 1 and 23 trajectories for Scene 2.

Then, **Fig. 14**, **Fig. 15**, and **Fig. 16** show the results of human detection. **Table 1** shows the maximum values of the precision, recall, and F-value for each method. For these cases, detection is done for 20 pedestrians.

From these results, it is evident that the performance of the human detector with the proposed priors is the best among the three detectors. When we employ uniform priors, the precision is low. This is because the detector only considers local image patterns and it cannot classify the difference between an actual person and other areas that have similar texture pattern. Although the priors from motion trajectories can be used to avoid such errors, this also reduces the recall rate because the distribution of the priors is too sparse for the observed scene.

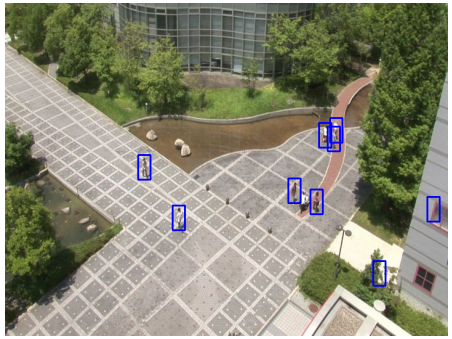


Fig. 14 Human detection: uniform priors.

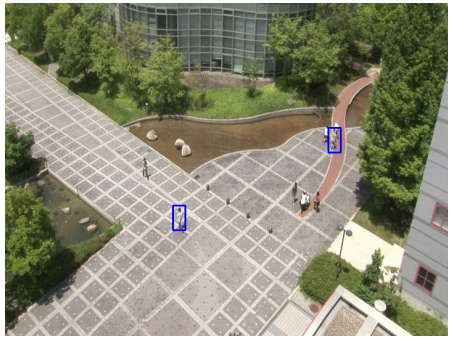


Fig. 15 Human detection: priors by trajectories.

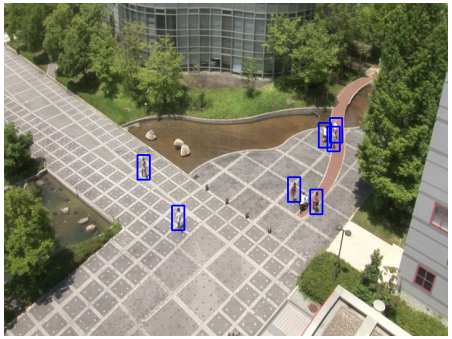


Fig. 16 Human detection: priors by trajectories and color.

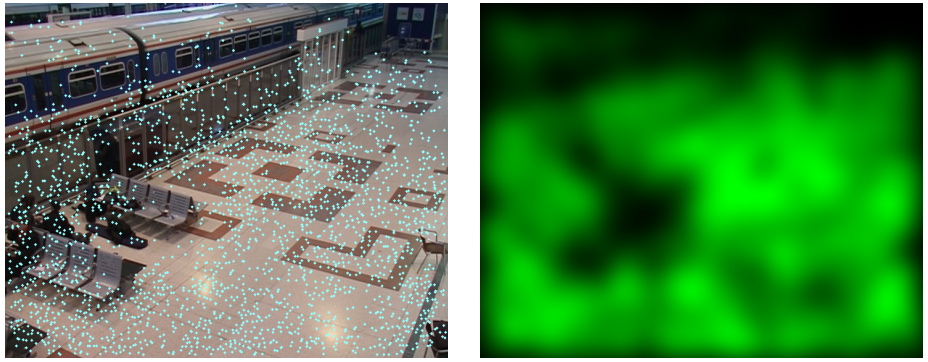
Table 1 Quantitative comparison of human detectors — Experiment 1.

Detector	Precision	Recall	F-value
Uniform Priors	0.63	0.83	0.72
Priors by Traj.	0.89	0.60	0.72
Priors by Traj. and Color	0.95	0.90	0.93

4.4 Experiment 2 — Estimation Results Using Automatically Detected Human Trajectories

The results shown in the previous section demonstrate the effectiveness of the proposed method. However, when we apply the method to actual application scenarios, it is impossible to avoid errors in human detection. In order to see the

Fig. 17 Scene 3: a sample of human detection results in Experiment 2.



(i) Particle Distribution (ii) Acquired Priors

Fig. 18 Scene 3: human existence priors by proposed method.

influence of errors on acquired priors, we applied the method to the scene shown in Fig. 7. This video consists of around 3,000 frames, and we applied the human detector to the video as discussed in Section 4.1. There would be more errors in human detection because of objects in the scene that have a similar appearance to actual humans. Actually, when we applied the human detector using a simple HOG feature and SVM classifier [6], there were some errors as shown in **Fig. 17**.

First, we applied the proposed method to 2,500 images of the data set. The obtained human existence priors are shown in **Fig. 18**. Using the priors, we carried out pedestrian detection for the remaining 500 images and compared the results with the other methods as in Section 4.3. **Table 2** shows the results. Note that we used 38 people walking on the road for learning and 12 people for testing.

Table 2 Quantitative comparison of human detectors for PETS 2006 data — Experiment 2.

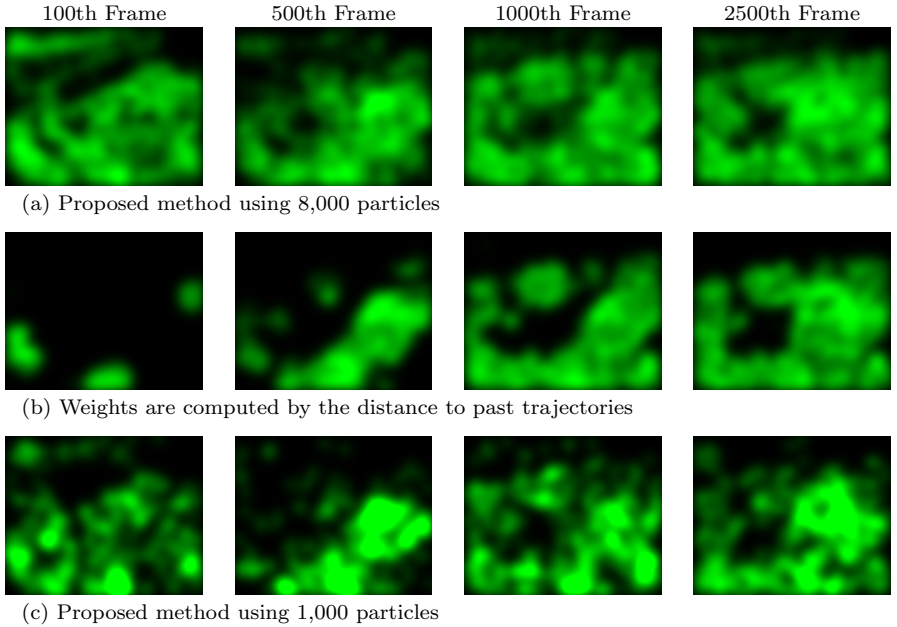
Detector	Precision	Recall	F-value
Uniform Priors	0.52	0.82	0.63
Priors by Traj.	0.67	0.68	0.68
Priors by Traj. and Color	0.77	0.74	0.76

Although the results are not as good as the previous experiment, we can see similar characteristics in the results. That is, the uniform priors give poor results, and considering solely human trajectories decreases the recall rate. In contrast, the proposed method has the best performance among the three methods. Note that this table shows the results which have the best F-value under different threshold values. If looking only at the difference in the recall ratios, the uniform priors seem to be the best among the three methods. However, this is not the case because the results of the uniform priors have the worst F-value, that is, many False-Positive samples appeared in the results.

The reason for the poor results of the proposed method is that no particle is generated at a position where no pedestrian has been detected since the beginning. For such regions the priors become small and a person appearing at the position is not detected unless a relatively large likelihood is given by the human detector from the definition of Eqs. (1) and (9). This cannot be avoided completely because of the nature of the proposed method. We would be able to reduce such False-Negative samples by adding a fixed value to the estimated prior and/or adjusting the threshold value. However, this is equivalent to discarding the past information and causes more False-Positive samples. This is a common issue when we employ the "memory" of past events for understanding current events. Incorporating other kinds of information, such as the semantic structure obtained in advance, would enable us to cope with this issue.

4.5 Comparing Time-varying Priors

Next, we show the detailed evaluation results in order to demonstrate the characteristics and appropriateness of the proposed method.



(a) Proposed method using 8,000 particles

(b) Weights are computed by the distance to past trajectories

(c) Proposed method using 1,000 particles

Fig. 19 Time-varying priors obtained by three methods for Scene 3.

As discussed in the previous sections, our method makes use of both past trajectories and color information in the observed images. First, we can see that this combination works adequately. The top row of **Fig. 19** shows the time-varying priors given by the proposed method for Scene 3, and the second row shows the priors obtained by the method that only utilizes past human trajectories, $\mathcal{L}^i_{color}(t) = 0$ in Eq. (8), for the same scene. As shown in Fig. 7, there is a wide passage in the image and high priors should be given there. By comparing these two results, it is evident that the proposed method yields approximately uniform priors on the passage using fewer images. This enables us to avoid the False-Negative errors discussed in the previous section.

The bottom row of the figure shows the results using 1,000 particles to represent the priors. As discussed at the beginning of Section 4, we used 8,000 particles in the previous experiments as well as in the top row of Fig. 19. The results show that using more particles produces more uniform priors. In this case, as discussed above, because the priors should be uniform, the results in the top row are better than the results in the bottom row.

From these results, we can see that considering both trajectories and color information makes it possible to estimate the priors efficiently. In addition we must use a high enough number of particles to estimate the priors correctly. From our observation, it is sufficient to use around 8,000 particles for the scenes used in the experiments. The appropriate number of particles depends on the area where pedestrians pass in an image.

5. Conclusion

In this study, we have proposed a method for acquiring the prior probability of human existence by using both past human trajectories and the color of an image. The proposed method is based on the assumption that a person can exist again in an area where he/she existed in the past. Through experiments, we confirmed that our proposed method can acquire the prior probability efficiently, and use it to realize highly accurate human detection.

As a future work, by incorporating sophisticated techniques which estimate the geometric structure discussed in Section 2, human existence priors could be estimated more accurately. As already discussed, we regarded the color at each pixel in an image as a fundamental feature representing the geometric structure, that is, regions having the same geometric properties have the same color. However, this assumption is not always satisfied. For example, a road is sometimes composed of two or more colors as shown in Fig. 7. In such cases, the proposed method does not estimate human existing priors correctly unless people are walking on all of the color regions. More sophisticated methods for estimating the geometric structure such as Ref. 5) are required in such cases.

In addition, it is possible and necessary to take into account higher information in order to represent and utilize the context of observing a scene. For example, while some works exploit a human motion model for predicting the current position of a person from its past trajectory [11], considering motion flow on the road as a context of the scene enables us to improve the accuracy of human tracking. However, when we incorporate higher contextual information such as motion flow,

it would become difficult to capture events which have never occurred. This is a common issue of "memory-based" methods such as Ref. 12) and the proposed one. We will explore solutions for this fundamental and challenging issue.

Acknowledgments The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. This work was partially supported by JSPS, Grant-in-Aid for Young Scientists (B) 19700166.

References

1) Leibe, B., Schindler, K. and Van Gool, L.: Coupled Detection and Trajectory Estimation for Multi-Object Tracking, *ICCV 2007*, pp.1–8 (2007).
2) Suzuki, T., Iwasaki, S., Kobayashi, Y., Sato, Y. and Sugimoto, A.: Incorporating environmental models for improving vision-based tracking of people, *Systems and Computers in Japan*, Vol.38, No.2, pp.1592–1600 (2007).
3) Hoiem, D., Efros, A.A. and Hebert, M.: Putting Objects in Perspective, *CVPR 2006* (2006).
4) Sugimura, D., Kobayashi, Y., Sato, Y. and Sugimoto, A.: Incorporating Long-Term Observations of Human Actions for Stable 3D People Tracking, *Proc. IEEE Workshop on Motion and Video Computing*, pp.1–7 (2008).
5) Hoiem, D., Efros, A.A. and Hebert, M.: Geometric Context from a Single Image, *ICCV 2005*, Vol.1, pp.654–661 (2005).
6) Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, *CVPR 2005*, Vol.II, pp.886–893 (2005).
7) Mount, D.M. and Arya, S.: ANN: A Library for Approximate Nearest Neighbor Searching. http://www.cs.umd.edu/~mount/ANN
8) Felzenszwalb, P.F. and Huttenlocher, D.P.: Efficient Graph-based Image Segmentation, *IJCV*, Vol.59, No.2, pp.167–187 (2004).
9) PETS 2006 Benchmark Data. http://www.cvg.cs.reading.ac.uk/PETS2006/data.html
10) Joachims, T.: SVMlight Support Vector Machine. http://svmlight.joachims.org/
11) Pellegrini, S., Ess, A., Schindler, K. and van Gool, L.: You'll never walk alone: Modeling social behavior for multi-target tracking, *ICCV 2009*, pp.261–268 (2009).
12) Mikami, D., Otsuka, K. and Yamato, J.: Memory-based particle filter for face pose tracking robust under complex dynamics, *CVPR 2009* (2009).

(Received November 10, 2009)
(Accepted August 6, 2010)
(Released November 10, 2010)

(Communicated by *Daisaku Arita*)

Hitoshi Habe was born in 1974. He received his B.E. and M.E. degrees in electrical engineering and D. Info. degree in intelligence science and technology from Kyoto University, Japan, in 1997, 1999, and 2006, respectively. After working for Mitsubishi Electric Corporation from 1999 to 2002, he joined Kyoto University as an assistant professor. Currently he is an assistant professor at Nara Institute of Science and Technology, Japan. From 2010 to 2011, he is visiting at the Department of Engineering, University of Cambridge, as a visiting researcher. His research interests include computer vision, pattern recognition, and image processing. He is a member of IPSJ, IEICE, and IEEE.

Hidehito Nakagawa received his B.E. degree from Osaka University in 2007 and his M.E. degree from Nara Institute of Science and Technology in 2009. He is currently working for KEYENCE Corporation.

Masatsugu Kidode is a professor at the Graduate School of Information Science, Nara Institute of Science and Technology (NAIST). He leads the Advanced Intelligence Laboratory at NAIST. After he received an M.S. degree in electrical engineering from Kyoto University in 1970, he worked for TOSHIBA Corp., where he developed a high-speed image processing system TOSPIX with several practical applications. Based on these research studies and technology developments, he received a Ph.D. degree in information engineering from Kyoto University in 1980. After 30-years of experience at a company, he joined NAIST in 2000. Since 2009, he has been serving as a vice president of NAIST. His main interests include intelligent media understanding techniques for real world applications such as home robots, human interface, intelligent systems, and so on. He has published over 150 scientific papers, technical articles and patents in image analysis and its practical applications. He has been honored by four Fellows from IEEE, IAPR, IPSJ, and IEICE.