Efficient Acquisition of Human Existence Priors from Motion Trajectories

Hitoshi Habe Hidehito Nakagawa Masatsugu Kidode
Graduate School of Information Science, Nara Institute of Science and Technology
8916–5, Takayama, Ikoma, Nara, JAPAN 630-0192
habe@computer.org

Abstract

This paper reports a method for acquiring the prior probability of human existence by using past human trajectories and the color of an image. The priors play important roles 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 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 it can realize highly accurate human detection using the obtained prior probability.

1. Introduction

In recent years, human beings have been increasingly subjected to video surveillance. Because manned observation is unfeasible, sophisticated techniques 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 video 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 give 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 the over- and miss-detection and improve the performance of human detectors and trackers[8, 4, 7].

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[8]. Some methods that estimate the geometric structure of a scene[3] can be used for accurate human detection[4]. Because the geometric structure of a scene directly affects human actions in a scene, we consider that 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 [7]. 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 human trajectories and color information of the images.

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 Figures 1 and 2, the motion trajectories will exhibit a sparse distribution on the road. Hence, in 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 is described above.

Figure 1. Scene 1

Figure 2. Scene 2

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 Exis tence Priors

Our main objective is to understand large-scale events occurring in the real worlds 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 proposed. The literatures 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 described 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 and roofs of buildings. If we can know such structures, i.e., *geometric structures*, in advance, it would 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 meanings 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 the priors.

While geometric structures have already been used for obtaining priors[8, 4], as discussed in Section 1, few studies have investigated the use of the semantic structure. 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.

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[1]. In other words, they do not consider information of other areas, such as the coocurrence with other objects and 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[8, 4, 7].

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

$$P(p|X) = \frac{P(X|p)P(p)}{P(x)} \tag{1}$$

where X 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 $P(X|p)$, which can be estimated by the human detectors, provide the posterior probability $P(p|X)$ 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.

3. Efficient Acquisition of Human Existence Priors

We describe our proposed method in this section.

3.1. Overview of Proposed Method

Figure 3 shows the process flow of the proposed method. Using the priors $P(p)$ at time $T-1$ and the observed image at time T, we first perform human detection ((1) in Figure3). Then, the priors are updated using the resultant human regions ((2) in Figure 3). These processes are iteratively conducted so that the priors adapt to variations in a scene. Note that our current implementation use 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 Parti cle Filter

Figure 4 shows how the priors are updated using the particle filter. Let $Y_t = \{y_1, y_2, \cdots, y_t\}$ denote a sequence of observed images from time 1 to t and T_t be

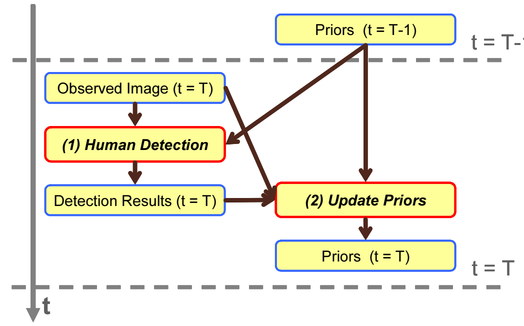

Figure 3. Efficient Acquisition of Human Existence Priors–Overview of Proposed Method

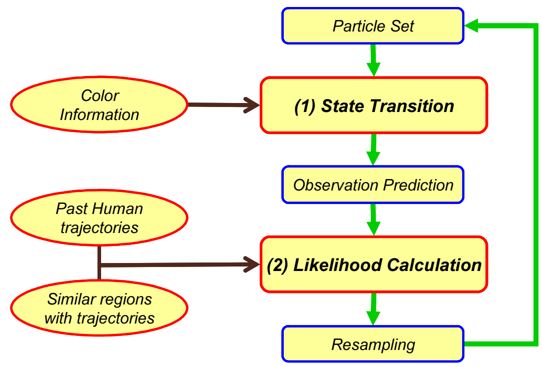

Figure 4. Updating Priors using Particle Filter

a group of observed human trajectories. Here, the prior at time t can be denoted as $P(p_t|Y_t, T_t)$. We represent it by the spatial distribution of a set of weighted samples $\{s^i_{t|t}|i = 1, 2, \ldots, N\}$ where $s^i_{t|t'}$ denotes the i-th particle at time t estimated from data until time t'. Let $\{\omega^i_t\}$ be weights for particle $\{s^i_t\}$. 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 $\{s^i_{t-1|t-1}\}$, we estimate the current set $\{s^i_{t|t-1}\}$. This estimation is performed using a state transition model $s^i_{t|t-1} = F(s^i_{t-1|t-1})$. The estimated sample set denotes the priors $P(p_t|Y_{t-1}, T_{t-1})$ that are derived from images and trajectories until time $t-1$.

(2) Computing Weight of Each Particle

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

(3) Re-sampling According to Ratios of Weights

Finally, we derive a particle set $\{s^i_{t|t}\}$ by re-sampling $\{s^i_{t|t-1}\}$ according to the ratios of weights $\{\omega^i_t\}$. The obtained $\{s^i_{t|t}\}$ represents the posterior probability distribution $P(p_t|Y_t, T_t)$ and it will be used as a prior at next time $t+1$.

In the above procedures, the state transition model $p_t = F(p_{t-1})$ makes use of the color information observed in an image. The weight function $\omega^i_t = H(s^i_{t|t-1})$ mainly employs past human motion trajectories. This is described in detail in the following section.

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 each particle at time $t-1$, which has coordinates $(x,\ y)$ as its state, to a position having a color similar to that of the current position. Figure 5 illustrates this process.

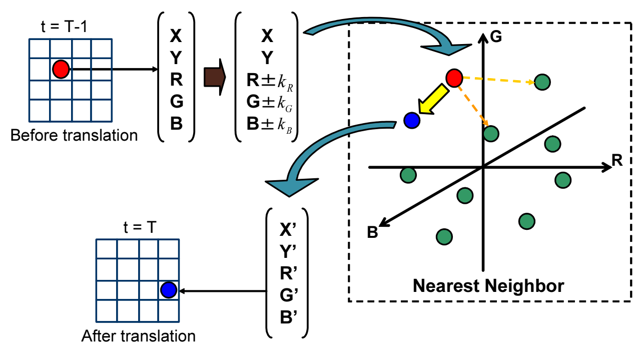

Figure 5. Estimating Current State using Color Information

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

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

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

where k_t is a 3D vector in which each component is a small number such as $k^{(j)}_t \in [-5,\ 5]$. Then, we find a pixel whose color is the same as or similar to that of c'^i_t by minimizing the L2-norm

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

When we find $c^i_t = \arg\min d^i$ for a particle at time t, approximate nearest neighbor (ANN) search[6] is applied for efficient computation. Finally, coordinates $(x,\ y)$ that are associated with c^i_t are selected as the current state of the particle.

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 exists; in other words, how people are likely to exist at a position 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 section.

Euclidean Distance to Past Trajectories A sequence of th coordinates of motion trajectories is denoted as follows:

$$\Psi_{trajectory} = \{(u^1_{tra}, v^1_{tra}), (u^2_{tra}, v^2_{tra}), \cdots (u^{N_{tra}}_{tra}, v^{N_{tra}}_{tra})\}, \tag{4}$$

where N_{tra} is the number of points on the trajectories. Let (u^i_t, v^i_t) be the coordinate of the i-th particle at time t. The minimum distance between the particle and the trajectories is given as

$$d^i_{Particle} = \min_{n=1,\ldots,N_{tra}} \sqrt{(u^i_t - u^n_{tra})^2 + (v^i_t - v^n_{tra})^2} \tag{5}$$

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

$$\mathcal{L}^i_{t\_dist} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{{d^i_{Particle}}^2}{2\sigma^2} \right) \tag{6}$$

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 time would be required to obtain adequate priors.

First, we segment an input image using the method described in [2]. 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. Figures 6 (1) and (2) show the integration process and the integrated region, respectively. Here, we assign a uniform

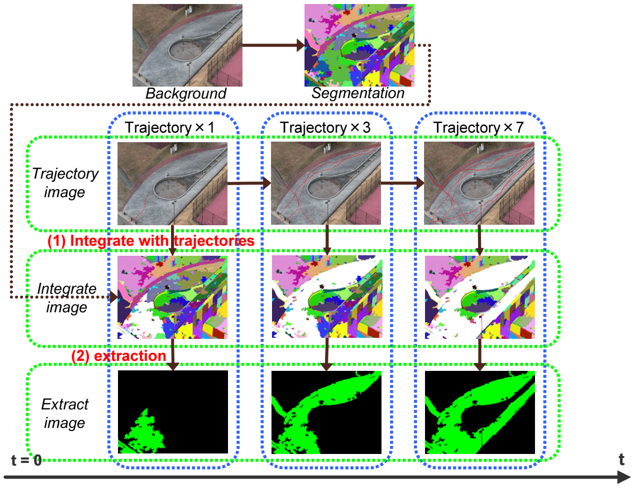

Figure 6. Similar Color Regions for Acquiring Priors

weight that has a high value when a particle lies in the integrated region, and a low weight when the particle lies outside the region. This weight function can be written as

$$d^i_{similar} = Z_t(u^i_t, v^i_t), \quad (7)$$

$$\mathcal{L}^i_{t\_similar} = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{{d^i_{similar}}^2}{2\sigma^2}\right), \quad (8)$$

where $Z_t()$ denotes the difference derived from the color at the particle; if it lies in the integrated region, it will be high, otherwise it will be small.

Now, we have two types of weight functions. Finally, these functions are integrated as given by Equation 9 and they are used for determining the weight of the i-th particle at time t.

$$\mathcal{L}^i_t = \mathcal{L}^i_{t\_dist} + \mathcal{L}^i_{t\_similar} \quad (9)$$

4. Experiments

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

4.1. Experimental Setup

We captured two outdoor videos for the experiments. Figures 1 and 2 show still images from these videos. In the experiments performed in this study, human motion trajectories are given manually. The trajectories are shown in Figures 7 and 9.

4.2. Acquiring Human Existence Priors

First, we will show the priors obtained by the proposed methods. By applying the proposed method for the trajectories, we obtain the priors, that is, the distribution of particles shown in Figures 8 and 10.

Figure 7. Scene 1: Human Motion Trajectories

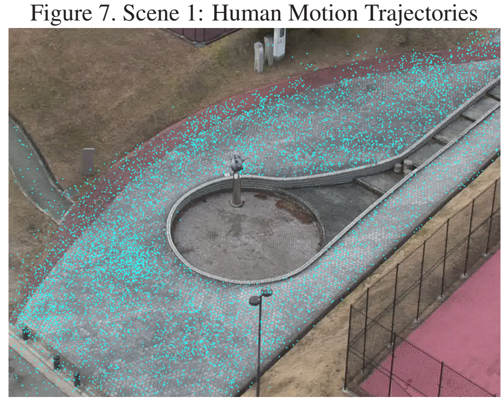

Figure 8. Scene 1: Particle Distribution

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.

4.3. Human Detection Incorporating Acquired Pri ors

Next, we show one of the applications of the obtained human existence priors. As discussed in Section 2.2, the priors can be applied to human detection and their use can improve its performance.

In order to confirm this, we perform human detection using a detector that consists of the HOG features[1] and the SVM classifier[5]. For quantitative comparison, we compute the precision, recall, and F-value, respectively defined as follows:

$$Precision = \frac{TP}{TP+FP}, \quad (10)$$

$$Recall = \frac{TP}{TP+FN}, \quad (11)$$

$$F = 2/\left(\frac{1}{Precision} + \frac{1}{Recall}\right), \quad (12)$$

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.

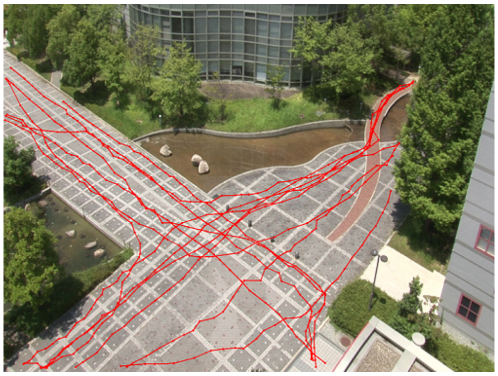

Figure 9. Scene 2: Human Motion Trajectories

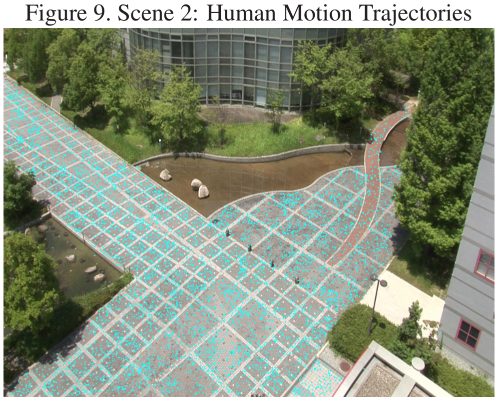

Figure 10. Scene 2: Particle Distribution

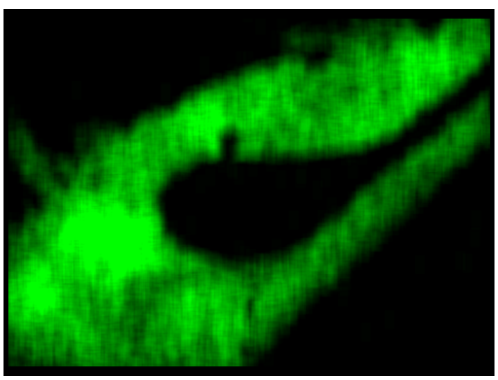

Figure 11. Scene 1: Human Existence Priors by Proposed Method

Figure 12. Scene 2: Human Existence Priors by Proposed Method

4.3.1 Human Detectors for Comparison

In this experiment, we compare the following three types of human detectors. These detectors differs in terms of the priors used for detection.

(1) Uniform Priors First, we perform human detection without considering priors. This is equivalent to using a uniform prior.

(2) Priors from Trajectories Second, 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 Finally, we use the proposed method that employs both past human trajectories and color information in an image. The priors are assigned using the method described in Section 3. The obtained priors are shown in Figures 11 and 12.

Figures 13, 14, and 15 show examples of detection results. Table 1 shows the maximum values of the precision, recal, and F-value for each method.

Table 1. Quantitative Comparison of Human Detectors

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

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.

5. Conclusion

In this study, we have proposed a method for acquiring the prior probability of human existence by using 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 it can realize

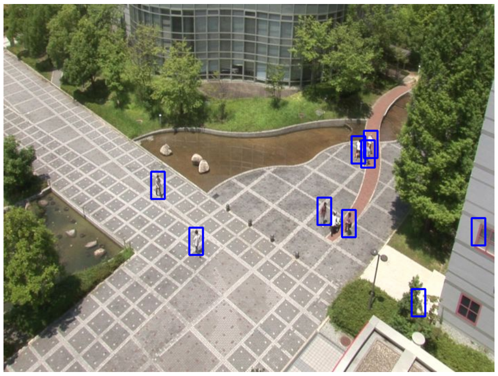

Figure 13. Human Detection: Uniform Priors

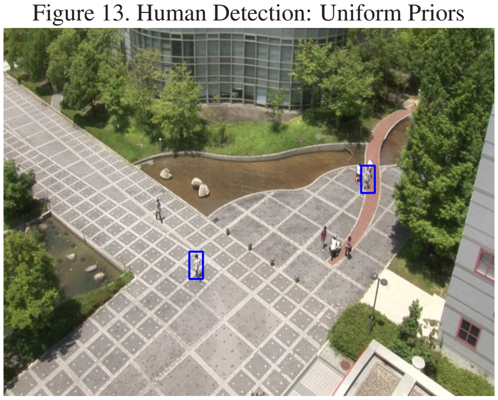

Figure 14. Human Detection: Priors by Trajectories

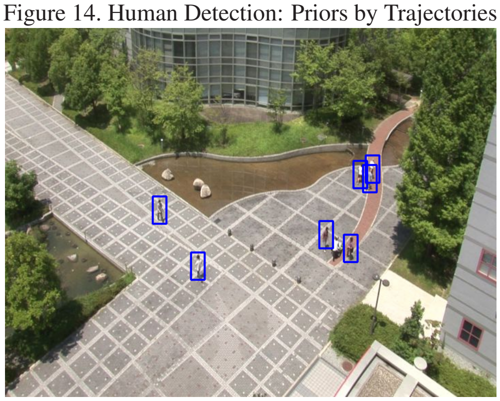

Figure 15. Human Detection: Priors by Trajectories and Color

highly accurate human detection using the obtained prior probability.

This work was partially supported by JSPS, Grant-in-Aid for Young Scientists (B) 19700166.

References

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of CVPR2005*, volume II, pages 886–893, 2005.

[2] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–187, 2004.

[3] D. Hoiem, A. A. Efros, and M. Hebert. Geometric context from a single image. In *IEEE International Conference on Computer Vision*, volume 1, pages 654–661, Los Alamitos, CA, USA, 2005. IEEE Computer Society.

[4] D. Hoiem, A. A. Efros, and M. Hebert. Putting objects in perspective. In *CVPR2006*, 2006.

[5] T. Joachims. Svmlight support vector machine. http://svmlight.joachims.org/.

[6] D. M. Mount and S. Arya. Ann: A library for approximate nearest neighbor searching. http://www.cs.umd.edu/ mount/ANN.

[7] D. Sugimura, Y. Kobayashi, Y. Sato, and A. Sugimoto. Incorporating long-term observations of human actions for stable 3d people tracking. In *Proc. IEEE Workshop on Motion and Video Computing*, pages 1–7, 2008.

[8] T. Suzuki, S. Iwasaki, Y. Kobayashi, Y. Sato, and A. Sugimoto. Incorporating environmental models for improving vision-based tracking of people. *Systems and Computers in Japan*, 38(2):1592–1600, 2007.