IPSJ Transactions on Computer Vision and Applications Vol.7 74–78 (July 2015)

[DOI: 10.2197/ipsjtcva.7.74]

Express Paper

Individuality-preserving Silhouette Extraction for Gait Recognition

Yasushi Makihara[1,a)] Takuya Tanoue[1,b)] Daigo Muramatsu[3,1,c)] Yasushi Yagi[1,d)]
Syunsuke Mori[2,e)] Yuzuko Utsumi[2,f)] Masakazu Iwamura[2,g)] Koichi Kise[2,h)]

Received: March 13, 2015, Accepted: April 20, 2015, Released: July 27, 2015

Abstract: Most gait recognition approaches rely on silhouette-based representations due to high recognition accuracy and computational efficiency, and a key problem for those approaches is how to accurately extract individuality-preserved silhouettes from real scenes, where foreground colors may be similar to background colors and the background is cluttered. We therefore propose a method of individuality-preserving silhouette extraction for gait recognition using standard gait models (SGMs) composed of clean silhouette sequences of a variety of training subjects as a shape prior. We firstly match the multiple SGMs to a background subtraction sequence of a test subject by dynamic programming and select the training subject whose SGM fit the test sequence the best. We then formulate our silhouette extraction problem in a well-established graph-cut segmentation framework while considering a balance between the observed test sequence and the matched SGM. More specifically, we define an energy function to be minimized by the following three terms: (1) a data term derived from the observed test sequence, (2) a smoothness term derived from spatio-temporally adjacent edges, and (3) a shape-prior term derived from the matched SGM. We demonstrate that the proposed method successfully extracts individuality-preserved silhouettes and improved gait recognition accuracy through experiments using 56 subjects.

Keywords: silhouette extraction, gait recognition, shape prior, graph-cut segmentation

1. Introduction

Person authentication from surveillance cameras play an increasingly important role in forensics (e.g., person re-identification and verification of a perpetrator and a suspect), and gait biometrics [13] have been considered as a promising cue for person authentication, since it can be utilized even if the perpetrator/suspect is captured at a distance from the surveillance camera.

Approaches to gait recognition mainly fall into two families: model-based and appearance-based ones. Among them, the appearance-based approaches have been dominant in the gait recognition community since they work well even for lower-resolution images with less computational cost than the model-based ones. In particular, a main stream of the appearance-based approaches exploit silhouette-based representations [5], [16], [19] since they are unaffected by clothing color and texture. Gait recognition accuracy using silhouette-based representations is, however, subject to silhouette quality.

1 Osaka University, Ibaraki, Osaka 567–0047, Japan
2 Osaka Prefecture University, Sakai, Osaka 599–8531, Japan
3 The National Institute of Information and Communications Technology, Osaka 530–0011, Japan
a) makihara@am.sanken.osaka-u.ac.jp
b) tanoue@am.sanken.osaka-u.ac.jp
c) muramatsu@nict.go.jp
d) yagi@am.sanken.osaka-u.ac.jp
e) mori_s@m.cs.osakafu-u.ac.jp
f) yuzuko@cs.osakafu-u.ac.jp
g) masa@cs.osakafu-u.ac.jp
h) kise@cs.osakafu-u.ac.jp

Silhouette extraction, i.e., foreground/background segmentation, has been studied for a long time in image processing and computer vision fields [1]. While traditional approaches to background subtraction exploit pixel-wise background modelling [8], recent approaches take adjacent connectivity or smoothness into consideration for better segmentation. A seminal work on this topic is graph-cut segmentation [2] and its variants: GrabCut [15] and mutual GrabCut [4]. In addition, soft segmentation a.k.a. alpha matte process of foreground/background, is also considered by the image segmentation community [9] and its effectiveness is demonstrated in the gait recognition community [6].

While these approaches work well when mis-assigned regions (e.g., over-segmentation in background and under-segmentation in foreground) are small enough to be corrected by imposing the smoothness, they do not work when they are too large to be corrected (e.g., the bulk of the under-segmentation in the leg region in **Fig. 1** (e)).

To solve this challenging task, a shape prior is incorporated in the segmentation framework [18]. For example, Liu and Sarkar [10] train an eigen stance, i.e., an eigen space of silhouettes at each gait stance, from clean silhouettes of multiple training subjects, and reconstruct silhouettes of a test subject through the eigen stance. The reconstructed silhouettes may, however, not preserve the individuality of the test subject since their variations are limited to the eigen space, i.e., a weighted linear sum of training subjects' silhouettes.

Wang et al. [20] also incorporate the shape prior in silhou-

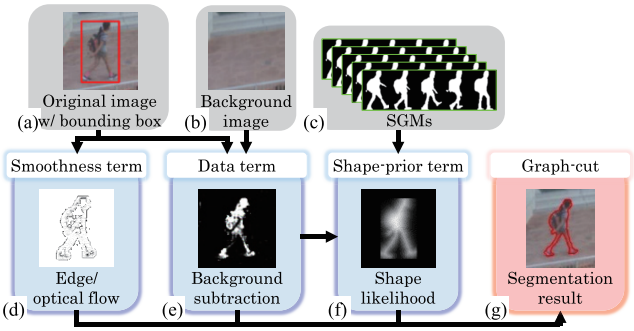


Fig. 1 Framework of the proposed method which leverages a shape prior (f) derived from multiple SGMs (c) as well as the data (e) and the smoothness (d) in graph-cut segmentation for better results (g).

ette extraction. They matched a standard gait model (SGM) to initially extracted silhouettes and improved the silhouettes while considering a balance between the matched SGM and the initial silhouettes containing the test subject's individuality in the graph-cut segmentation framework. However, since they use a single SGM, the improved silhouettes tend to be close to the single SGM and may reduce inter-subject variations.

We therefore propose a method of individuality-preserving silhouette extraction which efficiently exploits multiple SGMs in conjunction with the graph-cut segmentation framework. In this context, contributions of this paper are summarized as the following two points.

1. Individuality-preserving silhouette extraction using multiple SGMs: While previous studies [10], [20] may wash out the individuality in the silhouettes, our proposed method keeps the individuality as much as possible by selecting the best-fit SGM from multiple SGMs for each test subject and by balancing the matched SGM and the initial silhouettes containing the test subject's individuality in the graph-cut segmentation framework.

2. Accuracy improvement in gait recognition: While previous studies [10], [20] did not report the accuracy improvement in gait recognition, we demonstrate the proposed silhouette extraction actually improves gait recognition accuracy thanks to the individuality-preserving property described above.

2. Proposed Method

2.1 Problem Setting

In this study, we consider person authentication of pedestrians captured by two different cameras. Under this problem setting, we assume that the cameras are static and that background image sequences without pedestrians for background modelling are availble. Moreover, since we focus on silhouette extraction for gait recognition, we assume that bounding box sequences for individual pedestrians are given by well-established pedestrian detectors [3] and trackers [21].

2.2 Framework

As with most segmentation approaches, we adopt a graph-cut segmentation framework [2] which assigns a foreground/background label to each pixel through energy minimization. This is described in Fig. 1.

Given an original image and background images, we compute a foreground/background likelihood as a data term based on background subtraction, and a smoothness term to enhance foreground/background label consistency in the spatio-temporal proximity. In addition, we match multiple SGMs to the data term and compute the shape-prior term derived from the best-matched SGM. We then define an energy function $E(X)$ as a weighted linear sum of the data term $E_{dt}(X_q)$, the smoothness term $E_{sm}(X_p, X_q)$, and the shape-prior term $E_{sh}(X_q)$ as

$$E(X) = w_{dt}\sum_{q\in Q} E_{dt}(X_q) + w_{sm}\sum_{(p,q)\in P} E_{sm}(X_p, X_q) + w_{sh}\sum_{q\in Q} E_{sh}(X_q), \quad (1)$$

where Q and P are sets of sites (pixels) and edges (pairs of spatio-temporally adjacent pixels), X_q is a foreground/background label at the site q (FG: foreground, BG: background), X is a set of labels for all the sites Q, and w_{dt}, w_{sm}, and w_{sh} are weights to consider the tradeoff among individual terms. Finally, we obtain the optimal label assignment by minimizing the energy function $E(X)$ with the min-cut algorithm. We describe the details of the individual procedures in the following subsections.

2.3 Data Term

We briefly describe the data term in this subsection and refer the reader to Ref. [12] for more details.

We first train a pixel-wise background model as a single Gaussian from the given background image sequence, and compute the Mahalanobis distance $d_{bg,q}$ at each site q between an input pixel value and the trained background model. We then define the background data term $E_{dt}(X_q = BG)$ as

$$E_{dt}(X_q = BG) = \exp\left(-\kappa_{bg} d_{bg,q}\right), \quad (2)$$

where κ_{bg} is a hyper-parameter.

We subsequently train a foreground model as a Gaussian mixture model (GMM) based on the background subtracted result. We then compute Mahalanobis distances $d^k_{fg,q}$ between an input pixel value at each site q and the k-th component of the trained foreground GMM, and define the foreground data term $E_{dt}(X_q = FG)$ as

$$E_{dt}(X_q = FG) = \exp\left(-\kappa_{fg} \min_k d^k_{fg,q}\right), \quad (3)$$

where κ_{fg} is a hyper-parameter.

2.4 Smoothness Term

We first define a set of edges P as pairs of spatio-temporally adjacent pixels. While we simply use four connected neighbors for the spatial domain, we use optical flow correspondences [11] for the temporal domain. We then define the smoothness term as

$$E_{sm}(X_p, X_q) = \begin{cases} 0 & (X_p = X_q) \\ \exp\left(-\kappa_{sm}\frac{\|\boldsymbol{c}_q - \boldsymbol{c}_p\|^2}{\|\boldsymbol{c}_q + \boldsymbol{c}_p\|^2 + \varepsilon}\right) & (X_p \neq X_q), \end{cases} \quad (4)$$

where $\boldsymbol{c}_q$ is an RGB color vector at the site q, $\|\cdot\|$ stands for the L_2 norm, and κ_{sm} and ε are hyper-parameters.

2.5 Shape-prior Term

Firstly, we briefly describe matching between an input (i.e.,

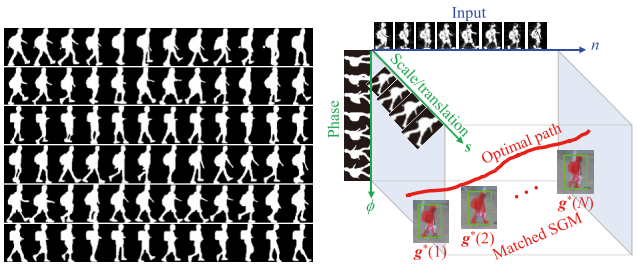


Fig. 2 SGMs (left) and its DP matching (right).

a background subtraction sequence) and an SGM. We refer the reader to Ref. [20] for more details.

We introduce a set of SGMs composed of a complete period of clean silhouette sequences from M training subjects as shown in **Fig. 2** (left). Since the given bounding box sequences may contain small deviations from the ground truth, we consider variations to scaling s, horizontal translating s_x, and vertical translating s_y (let a set of them be $\boldsymbol{s} = [s, s_x, s_y]^T$) as well as phase ϕ (i.e., gait stance) in the following matching procedure.

Once we define the input background subtraction sequence cropped within the bounding box sequence as $\{\boldsymbol{f}(n)\}$ ($n = 1, \ldots, N$), where N is the number of frames, and the variations of the SGMs as $\{\boldsymbol{g}_m(\phi, \boldsymbol{s})\}$ ($m = 1, \ldots, M$), we can obtain a sequence of matched SGMs for the m-th training subject by dynamic programming (DP) as $\{\boldsymbol{g}_m(\phi^*_m(n), \boldsymbol{s}^*_m(n))\}$, where $\{\phi^*_m(n)\}$ and $\{\boldsymbol{s}^*_m(n)\}$ ($n = 1, \ldots, N$) are the sets of the phase and the scale/transition along the optimal path of the DP matching to the m-th SGM (see Fig. 2 (right)), respectively. Moreover, we denote the optimal cumulative cost for the m-th SGM C^*_m, by simply choosing the training subject with the minimal DP matching cost as $m^* = \arg\min_m C^*_m$.

Subsequently, we formulate the shape-prior term based on the matched SGM $\{\boldsymbol{g}_{m^*}(\phi^*_{m^*}(n), \boldsymbol{s}^*_{m^*}(n))\}$ (we denote it $\{\boldsymbol{g}^*(n)\}$ for simplicity). After we compute the signed distance $d_{sh,q}$ of the matched SGM $\{\boldsymbol{g}^*(n)\}$ for the site q (i.e., positive and negative values for inside and outside of the silhouette, respectively), we compute the background/foreground shape-prior terms using a sigmoid function as

$$E_{sh}(X_q = BG) = \frac{1}{1 + \exp(-\kappa_{sh} d_{sh})} \quad (5)$$

$$E_{sh}(X_q = FG) = 1 - E_{sh}(X_q = BG), \quad (6)$$

where κ_{sh} is the gain for this sigmoid function.

A property of this representation is that the shape-prior is mitigated near the silhouette contour while it becomes stronger as the site is further from the silhouette contour (i.e., probable inside or outside, see Fig. 1 (f)). Thanks to this property, we avoid making the segmentation results too much close to the matched SGM, which is beneficial when the matched SGM is deviated from the ground truth of a test subject. Moreover, thanks to the multiple SGMs, we can suppress this deviation by selecting the best matched SGM. We therefore successfully handle the tradeoff between the data and the shape prior.

3. Experiments

3.1 Setup

We captured image sequences of 56 subjects (pupils) from two

(a) Camera 1 (b) Camera 2

Fig. 3 Example images for each camera.

network cameras (let them be Camera 1 and Camera 2, respectively) installed in an elementary school (see **Fig. 3**)[*1] and each image sequence contained 25 to 35 frames. We then divided the 56 subjects into 6 training subjects and the other 50 test subjects for evaluating silhouette extraction and gait-based person authentication, where Camera 1 and Camera 2 were used as enrollment (gallery) and query (probe), respectively. We annotated clean silhouette sequences manually to construct the SGMs for the training subjects, and annotated bounding box sequences for the test subjects.

We compared the proposed method with five benchmarks: graph-cut [2], GrabCut [15], mutual GrabCut [4][*2], eigen stance [10][*3], and graph-cut with a single SGM [20][*4]. We experimentally set the hyper-parameters: tradeoff of individual terms as $w_{dt} = 0.7$, $w_{sm} = 1.0$, $w_{sh} = 0.3$, the data term as $\kappa_{bg} = 0.02$ (0.01 for Camera 2) and $\kappa_{fg} = 0.3$, the smoothness term as $\kappa_{sm} = 0.01$, $\varepsilon = 63$, and the shape-prior term as $\kappa_{sh} = 0.2$. As for graph-cut [2], we experimentally set the tradeoff of individual terms as $w_{dt} = 1.0$ and $w_{sm} = 1.0$.

3.2 Evaluation on Silhouette Extraction

We first show typical results of silhouette extraction in **Fig. 4** as a qualitative evaluation. As a result, we can see that the benchmarks without the shape priors (Fig. 4 (b)(c)(d)) suffer from over-segmentation of the shadow cast under the leg regions or under-segmentation of the leg regions.

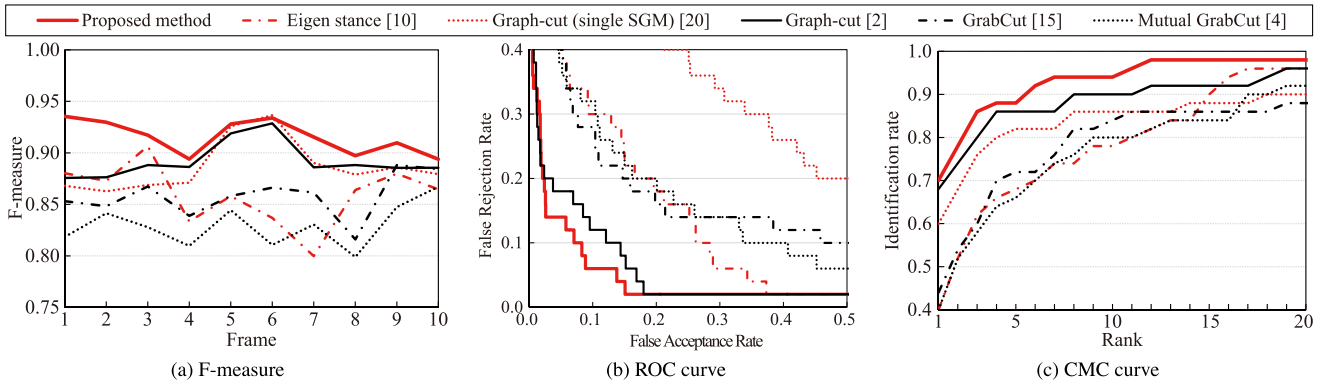
As for the benchmarks with the shape prior, the method [10] basically limits the reconstructed silhouette to the eigen stance and also substitutes an input background subtraction value for each pixel when the reconstructed silhouette value is uncertain, i.e., neither probable foreground nor background. The obtained silhouette is therefore still noisy (Fig. 4 (e)). Moreover, the method [20] relies on a single SGM, the extracted silhouette is attracted to the matched single SGM (Fig. 4 (f)), which is deviated from the ground truth and loses the individuality as a result. On the other hand, the proposed method successfully obtained the individuality-preserving silhouette (Fig. 4 (g)), which is similar to

[*1] We obtained approval for the use of the captured images with the network cameras for research purpose from the director and parent-teacher association of the elementary school.

[*2] We partly used OpenCV's implementation of GrabCut for Refs. [4], [15] and set a 10-pixels marginal region outside of the bounding box as a background sampling region while we left the other parameters as default.

[*3] Although we used the DP matching for phase (gait stance) estimation instead of a population hidden Markov model used in the original paper [10], we confirmed that the estimated phases are accurate enough for fair comparison.

[*4] The single SGM was experimentally chosen from the six training subject.



(a) F-measure (b) ROC curve (c) CMC curve

Fig. 5 Quantitative evaluation on silhouette extraction and gait recognition.

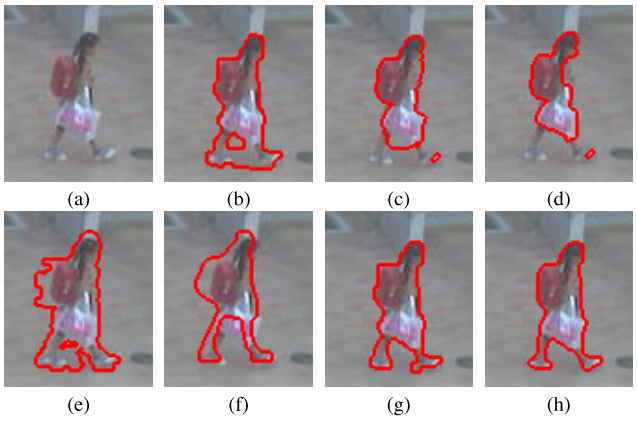


Fig. 4 Results of silhouette extraction (red lines: silhouette contour). (a) Original image, (b) Graph-cut [2], (c) GrabCut [15], (d) Mutual GrabCut [4], (e) Eigen stance [10], (f) Graph-cut w/ single SGM [20], (g) Proposed method, (h) Ground truth.

the ground truth (Fig. 4 (h)).

We also show F-measures [14] of the segmentation results for the first 10 frames in the test image sequence as a quantitative evaluation in **Fig. 5** (a). As a result, we confirmed that the proposed method outperformed the benchmarks for each frame.

3.3 Evaluation on Gait Recognition

We then evaluated the effectiveness of the proposed silhouette extraction in gait recognition, i.e., gait-based person authentication. For this purpose, we adopted GEI [5] as the most widely used silhouette-based gait feature, and matched them by Euclidean distance for simplicity.

The gait recognition accuracy was evaluated by a receiver operating characteristics (ROC) curve for a verification scenario (i.e., one-to-one matching) and a cumulative matching characteristics (CMC) curve for an identification scenario (i.e., one-to-many matching) as shown in Fig. 5 (b) and (c). As a result, the proposed method achieved the lowest error rates in the ROC curve and also the best identification rates over all the ranks in the CMC curve, because it kept the individualities in the extracted silhouette by using multiple SGMs. In other words, it avoids the problem with the extracted silhouette being too close to a single SGM.

4. Conclusion

We proposed a method of silhouette extraction for gait recognition. We formulated the proposed method in a graph-cut segmentation framework and incorporated the matching results of the SGMs as the shape-prior term. In addition, we exploited multiple SGMs in order to represent the individualities of the test subjects, which led to improvements in silhouette extraction as well as gait recognition.

Since the proposed method requires a relatively high computational cost through DP matching of the multiple SGMs under the multi-dimensional search space (i.e., phase, scale, and translations), a future avenue of research involves improvement in computational efficiency using analytical DP [17] in conjunction with fast approximate nearest neighbor search [7].

Acknowledgments This work was partly supported by a JSPS Grant-in-Aid for Scientific Research (A) 15H01693, "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society."

References

[1] Bouwmans, T., Porikli, F., Horferlin, B. and Vacavant, A.: *Background Modeling and Foreground Detection for Video Surveillance: Traditional and Recent Approaches, Implementations, Benchmarking and Evaluation*, CRC Press, Taylor and Francis Group (2014).
[2] Boykov, Y. and Funka-Lea, G.: Graph Cuts and Efficient N-D Image Segmentation, *Int. J. Comput. Vision*, Vol.70, No.2, pp.109–131 (2006).
[3] Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, *Prof. 18th IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Vol.2, pp.886–893 (2005).
[4] Gao, Z., Shi, P., Karimi, H. and Pei, Z.: A mutual GrabCut method to solve co-segmentation, *EURASIP J. Image and Video Processing*, Vol.2013, No.1, pp.1–11 (2013).
[5] Han, J. and Bhanu, B.: Individual Recognition Using Gait Energy Image, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.28, No.2, pp.316–322 (2006).
[6] Hofmann, M., Schmidt, S.M., Rajagopalan, A. and Rigoll, G.: Combined Face and Gait Recognition using Alpha Matte Preprocessing, *Proc. 5th IAPR Int. Conf. Biometrics*, New Delhi, India, pp.1–8 (2012).
[7] Iwamura, M., Sato, T. and Kise, K.: What Is the Most Efficient Way to Select Nearest Neighbor Candidates for Fast Approximate Nearest Neighbor Search?, *Proc. 14th International Conference on Computer Vision (ICCV 2013)*, pp.3535–3542 (2013).
[8] Lee, D.-S.: Effective Gaussian Mixture Learning for Video Background Subtraction, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.27, No.5, pp.827–832 (2005).
[9] Levin, A., Lischinski, D. and Weiss, Y.: A Closed-Form Solution to Natural Image Matting, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.30, No.2, pp.228–242 (2008).
[10] Liu, Z. and Sarkar, S.: Effect of Silhouette Quality on Hard Problems in Gait Recognition, *IEEE Trans. Systems, Man, and Cybernetics Part B: Cybernetics*, Vol.35, No.2, pp.170–183 (2005).
[11] Lucas, B. and Kanade, T.: An iterative image registration technique with an application to stereo vision, *Proc. 7th Int. Joint Conf. Artificial Intelligence*, pp.674–679 (1981).

[12] Makihara, Y. and Yagi, Y.: Silhouette Extraction Based on Iterative Spatio-Temporal Local Color Transformation and Graph-Cut Segmentation, *Proc. 19th Int. Conf. Pattern Recognition*, Tampa, Florida USA, pp.1–4 (2008).
[13] Nixon, M.S., Tan, T.N. and Chellappa, R.: *Human Identification Based on Gait*, Int. Series on Biometrics, Springer-Verlag (2005).
[14] Powers, D.M.W.: Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation, *Journal of Machine Learning Technologies*, Vol.2, No.1, pp.37–63 (2011).
[15] Rother, C., Kolmogorov, V. and Blake, A.: GrabCut -Interactive Foreground Extraction using Iterated Graph Cuts, *ACM Trans. Graphics*, Vol.23, No.3, pp.309–314 (2004).
[16] Sarkar, S., Phillips, J., Liu, Z., Vega, I., ther, P.G. and Bowyer, K.: The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.27, No.2, pp.162–177 (2005).
[17] Uchida, S., Fujimura, I., Kawano, H. and Feng, Y.: Analytical Dynamic Programming Tracker, *Proc. 10th Asian Conf. Computer Vision*, pp.296–309, Springer Berlin Heidelberg (2010).
[18] Vu, N. and Manjunath, B.: Shape prior segmentation of multiple objects with graph cuts, *Proc. 21th IEEE Conf. Computer Vision and Pattern Recognition*, pp.1–8 (2008).
[19] Wang, C., Zhang, J., Wang, L., Pu, J. and Yuan, X.: Human Identification Using Temporal Information Preserving Gait Template, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.34, No.11, pp.2164–2176 (2012).
[20] Wang, J., Makihara, Y. and Yagi, Y.: Human Tracking and Segmentation Supported by Silhouette-based Gait Recognition, *Proc. 2008 IEEE Int. Conf. Robotics and Automation*, pp.1698–1703 (2008).
[21] Wang, J. and Yagi, Y.: Integrating Color and Shape-texture Features for Adaptive Real-time Object Tracking, *IEEE Trans. Image Processing*, Vol.17, No.2, pp.235–240 (2008).

(Communicated by *Yoshito Mekada*)