Automatic Person Identification using Multiple Cues

Danuwat Swangpol and Thanarat Chalidabhongse

Faculty of Information Technology, King Mongkut s Institute of Technology Ladkrabang, Bangkok, Thailand
(Tel: +66-2-737-2551; E-mail: s4067410@kmitl.ac.th , thanarat@it.kmitl.ac.th)

Abstract: This paper describes a method for vision-based person identification that can detect, track, and recognize person from video using multiple cues: height and dressing colors. The method does not require constrained target s pose or fully frontal face image to identify the person. First, the system, which is connected to a pan-tilt-zoom camera, detects target using motion detection and human cardboard model. The system keeps tracking the moving target while it is trying to identify whether it is a human and identify who it is among the registered persons in the database. To segment the moving target from the background scene, we employ a version of background subtraction technique and some spatial filtering. Once the target is segmented, we then align the target with the generic human cardboard model to verify whether the detected target is a human. If the target is identified as a human, the card board model is also used to segment the body parts to obtain some salient features such as head, torso, and legs. The whole body silhouette is also analyzed to obtain the target s shape information such as height and slimness. We then use these multiple cues (at present, we uses shirt color, trousers color, and body height) to recognize the target using a supervised self-organization process. We preliminary tested the system on a set of 5 subjects with multiple clothes. The recognition rate is 100% if the person is wearing the clothes that were learned before. In case a person wears new dresses the system fail to identify. This means height is not enough to classify persons. We plan to extend the work by adding more cues such as skin color, and face recognition by utilizing the zoom capability of the camera to obtain high resolution view of face; then, evaluate the system with more subjects.

Keywords: Person Identification, Active Camera, Pattern Recognition, Image Analysis.

1. INTRODUCTION

As the need of intelligent system for security and surveillance grows rapidly, it becomes more urgent and important to establish viable methods for automatic person identification that can detect, track and recognize persons in real-time. Our goal is to develop a person identification system that can detect, track, and recognize people from video using multiple cues such as face, shape, clothes, skin color and gender. The system must not require constrained target s pose or fully frontal face image to identify the person. Our underlying motivation is to develop an automatic system to guard a watched area such as offices, dormitories, apartments, etc. The system is responsible to detect target that approaches the watched area, then identify the target. If the target is known person, the system put the recognition result and time stamp on the entry log file. Otherwise, it sends alarm signal to an authorized person via instant message over the network or SMS message over the mobile phone.

The main focus of this paper is on the person identification from unconstrained video part. First, the system, which is connected to a pan-tilt-zoom camera, detects target using motion detection and human cardboard model. The system keeps tracking the moving target while it is trying to identify whether it is a human and identify who it is among the registered persons in the database. This continuous identifying while tracking can provide a significant benefit as mentioned in [1] as *continuity of identity*. To segment the moving target from the background scene, we employ a version of background subtraction technique [2] and some spatial filtering. Once the target is segmented, we then align the target with the generic human cardboard model to verify whether the detected target is a human. If the target is identified as a human, the card board model is also used to segment the body parts to obtain some salient features such as head, torso, and legs. The whole body silhouette is also analyzed to obtain the target s shape information such as height and slimness. We then use these multiple cues (at present, we uses shirt color, trousers color, and body height) to recognize the target using a supervised self-organization process.

The organization of this paper is as follows: Section 2 reviews the literature related to our work. Section 3 describes our proposed methods in segmenting and identifying person from images. Section 4 presents the experiments and results. Finally, we conclude the work and discuss about future works in Section 5.

2. LITERATURE SURVE

In this section, we review existing techniques relevant to the problem of detecting, tracking, and identifying people in video. We divide into 3 subsections: first reviews person identification techniques, second reviews methods in detecting and tracking human in video, and the last subsection discusses about works that tracking and identifying human using active camera(s).

2.1 Person Identification

As one of the most successful applications of image analysis and understanding, person identification in video has recently received significant attention, especially during the past several years. To identify a person, most of the early works focused on biometrics such as fingerprint, iris, and face recognition. However, fingerprint and iris require the subject to directly interact with the sensors. This limits the domain of applications. Face recognition seems to be more feasible in the sense that it uses passive sensor. For exhaustive surveys of face analysis techniques, the reader is referred to ang et al. [3] for survey on face detection and Zhao et al. [4] for face recognition survey. However, most of the typical systems work on static high-resolution frontal face images. These are not feasible to some application domain such as surveillance. In surveillance, people might not need to know they are observed, and cameras are usually placed in distance from the subjects. In this circumstance, obtaining high-resolution frontal face images are very difficult. There are typically two approaches solving this

problem. One uses other cues such as shape and gait [5,6] to recognizing person, the other trying to model the face in multiple views [7-9].

2.2 uman Detection and Tracking

Human detection and tracking have been very active area of research in the past two decades. Two comprehensive surveys of computer vision-based human motion analysis have been published by Gavrila in 1999 [10], and by Moeslund and Granum in 2001 [11]. More recently, Kentaro et al. [12] presented a new exemplar-based probabilistic paradigm for visual tracking. Their approach, called Metric Mixture (M^2), combines the advantages of exemplar-based models [13] with a probabilistic framework introduced in [14] into a single probabilistic exemplar-based paradigm. Sullivan and Carlsson [15] presented a method in recognizing and tracking human in action. In their approach, view-based activity recognition serves as an input to a human body location tracker. By recognizing the image of a person s posture as corresponding to a particular stored key frame, they then map body locations from the key frames to actual frames using shape matching algorithm based on appearance similarity.

2.3 uman Tracking using Active Cameras

Recently, many researchers have been working on active and multiple cooperative cameras. This approach takes benefit of multi-scale imaging. Using wide-angle or stereo cameras to segment and track target in video, then use active camera(s) to zoom-in to capture face region for identification. The works in this category include work done by Peixoto et al. [16]. They use a wide-angle camera plus a ground plane assumption to estimate 3D location of the object. Then, the object is tracked using a binocular active camera. Another work done by Collins et al. [17] uses multiple cooperative cameras to detect, track, and recognize person. With multi-view images from multiple calibrated cameras, they can estimate 3D location of the target using triangulation and groundplane assumption. The closest related work in the literature we found is the one proposed by Hampapur et al. [1]. Their goal is to build a system that can answer the who is where question. Their system uses two static cameras for wide baseline stereo to estimate 3D position of the subject s head using triangulation technique. Then another set of two Pan-Tilt-Zoom (PTZ) cameras is used to zoom in on the moving target and catalog the face. However, they did not mentioned in detail about their method in identifying.

3. OUR APPROAC

Our approach use only one PTZ camera connected to the PC. The camera is mounted and points to the entrance area of the lab. Once it detects motion, it will keep tracking while identifying the moving object. Fig. 1 shows a block diagram of our system. The processes are divided into three phases: initialization, segmentation, and identification phases.

3.1 Initiali ation Phase

In this phase, personal models and background models are constructed. To build background model, the camera is panned in certain step while capturing background scene; a simple panorama background model is then constructed. Fig.2 shows the background model. In this work, background modeling and background subtraction technique proposed by Horprasert et al. [2] is used.

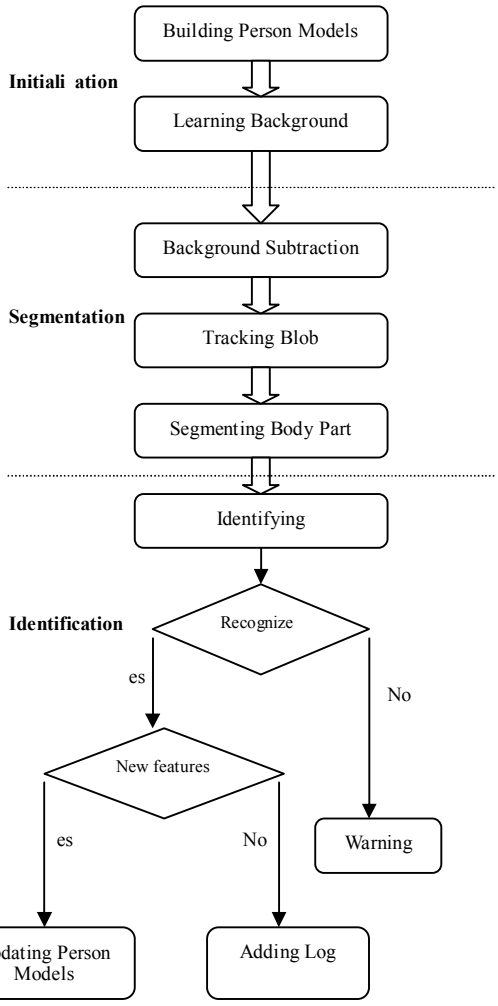

Fig. 1 Block diagram of our system.

To construct initial person models, we capture video sequence of each subject while walking into an observed area. The system then subtracts the input video from the background model. Result of the subtraction is shown in Fig.3. Then, the cardboard model is used to segment body parts (details in next subsection). The color of torso, color of legs, and height are then registered as feature of the person. We then iterate the process, until we register all the known subjects.

3.2 Segmentation Phase

This phase concerns with detecting and segmenting the target. At running mode, the camera is initially set to watch the entrance area of the room. Whenever, there is a person come in the scene, the system detects it using background subtraction. A spatial median filtering is employed to clean some erroneous segmentation. Connected component analysis is then

Fig. 2 Background model.

Fig. 3 Subtraction results.

performed. Only big regions remain. A simple cardboard model [18] is then used to verify whether the segmented region is person-like. While identifying, the camera keeps tracking the subject with Pan-Tilt capability. This enhances the identifying efficiency. The cardboard is also used to segment salient body parts such as head, torso, and legs. Subject s height, torso color, and leg color are then computed, taken as a feature vector for this subject. Fig.4 shows a result of body part segmentation.

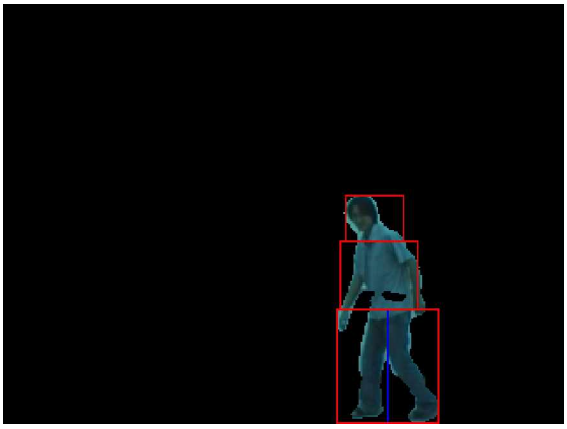

Fig. 4 Body part segmentation result.

3.3 Identification Phase

To recognize the subject, we utilize feedforward backpropagation neural networks. At current version, feature vector composes of 7 values: H, T_R, T_G, T_B, L_R, L_G, L_B, where H, T, and L are height, torso, and leg orderly. R, G, B are typical color dimension. The networks compose of 3 layers with 7 input nodes which correspond to the feature vector. The hidden layer contains hidden nodes that are fully linked to the input and output nodes. The number of output nodes are $N + 1$, where N is the number of register persons. The other node is for unknown person answer.

4. DISCUSSION

The preliminary system was implemented using C/C++ and OpenCV library running on a Pentium4 PC. The camera is Sony EVI-D100 connected to the PC via a frame grabber. To evaluate the system, we tested it on a set of 5 subjects with multiple clothes. Fig. 5 shows some of the training set of data and the segmentation results. The recognition rate is 100% if the person is wearing the clothes that were learned before. In case a person wears new dresses the system fail to identify. This means height is not enough to classify persons. Other available collateral information that we can extract from image such as skin color, gender, face, or speech may be used in enhancing recognition.

We plan to extend the work as follows:

- § Working on more sophisticate background subtraction method from moving camera video sequence. Example of works in the literature include Sugaya and Kanatani [18], and Bartoli et al. [19].
- § Adding more cues; skin color and face recognition by utilizing the zoom capability of the camera to obtain high-resolution view of face.
- § Evaluating the system with more subjects.

REFERENCES

[1] A. Hampapur, S. Pankanti, A. Senior, .L.Tian, L. Brown, and R. Boll, Face Cataloger: Multi-Scale Imaging for Relating identity to Location , Proc. IEEE Intl . Conf. Advanced Video and Signal Based Surveillance (AVSS 2003), IEEE Press, 2003, pp.13-20.

[2] T. Horprasert, D. Harwood, and L. S. Davis, A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection , Proc. IEEE Frame Rate Workshop, Greece, 1999.

[3] MH. ang, D.J.Kriegman, and N. Ahuja, Detecting Faces in Images: A Survey , IEEE Tran. Pattern Analysis and Machine Intelligence (PAMI), 24(1): 34-58, 2002.

[4] W . Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: A literature survey , ACM Comput. Survey, 35(4): 399-458, 2003.

[5] A. Kale, A. K. R. Chowdhury, and R. Chellappa, Towards a View Invariant Gait Recognition Algorithm , dddd

[6] S. Sarkar, P.J. Phillips, Z. Liu, I. Robledo, P. Grother, and K. Bowyer, The Human ID Gait Challenge Problem: Data Sets, Performance, and Analysis , IEEE Transaction on Pattern Analysis and Machine Intelligence, 27(2): 162-177, 2005.

[7] . Li, S. Gong, and H. Liddell, Support Vector Regression and Classification Based Multi-view Face Detection and Recognition , Proc. Intl . Conf. Face and Gesture Recognition (FG 2000), France, 2000.

[8] C. Sanderson, and S. Bengio, Extrapolating Single View Face Models for Multi-View Recognition , Proc. Intl . Conf. Intelligent Sensors, Sensor Networks, and Information Processing, Australia, 2004.

[9] . Gao, S.C. Hui, and A.C.M. Fong, A Multiview Facial Analysis Technique for Identity Authentication , IEEE Pervasive Computing, IEEE Communication Society, 2003.

[10] D.M. Gavrila, The Visual Analysis of Human Movement: A Survey , Computer Vision and Image Understanding (CVIU), 1999.

[11] T.B. Moeslund, and E. Granum, A Survey of Computer Vision-based Human Motion Capture , Computer Vision and Image Understanding (CVIU), 2001.

[12] K.Toyama, and A. Blake, Probabilistic Tracking with Exemplars in a Metric Space , Intl . Journal of Computer Vision, 48(1): 9-19, 2002.
[13] D. Gavrila, and V. Philomin, Real-time Object Detection for Smart Vehicles , Proc. Intl . Conf. Computer Vision, pp:87-93, 1999.
[14] B. Frey and N. Jojic, Learning Graphical Models of Images, Videos, and their Spatial Transformations , Proc. Conf. Uncertainty in Artificial Intelligence, 2000.
[15] J. Sullivan and S. Carlsson, Recognizing and Tracking Human Action , Proc. European Conf. Computer Vision, 2002.
[16] Peixoto, Batista, and Araujo, A Surveillance System Combining Peripheral and Foveated Motion Tracking , Proc. Intl . Conf. Pattern Recognition, 1998.
[17] R.T. Collins, A.J. Lipton, H. Fujiyoshi, and T. Kanade, Algorithms for Cooperative Multisensor Surveillance , Proceedings of the IEEE, 89(10):1456-1477, 2001.
[18] . Sugaya and K. Kanatani, Extracting Moving Objects from a Moving Camera Video Sequence , Proc. 10th Symposium on Sensing via Image Information (SSII 2004), Japan, 2004, pp:279-284.
[19] A. Bartoli, N. Dalal, B. Bose, and R. Horaud, From Video Sequence to Motion Panoramas , Proc. the Workshop on Motion and Video Computing (MOTION 02), IEEE Press, 2002.

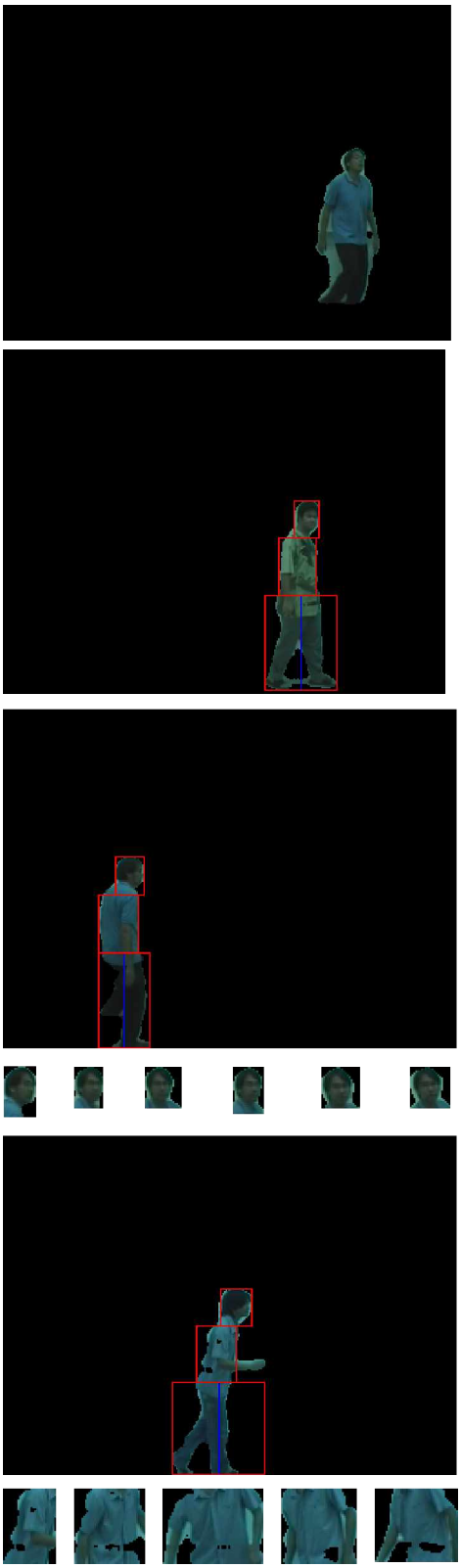

Fig. 5 More segmentation result images