IDENTIFICATION AND RECONSTRUCTION OF COMPLETE GAIT CYCLES FOR PERSON IDENTIFICATION IN CROWDED SCENES

Martin Hofmann, Daniel Wolf and Gerhard Rigoll
Institute for Human-Machine Communication, Technische Universität München
Arcisstr. 21, Munich, Germany

Keywords: Gait recognition, Person identification, Occlusions, Gait cycle reconstruction, Graph cuts.

Abstract: This paper addresses the problem of gait recognition in the presence of occlusions. Recognition of people using their gait has been an active research area and many successful algorithms have been presented. However to this point non of the methods addresses the problem of occlusion. Most of the current algorithms need a full gait cycle for recognition. In this paper we present a scheme for reconstruction of full gait cycles, which can be used as preprocessing step for any state-of-the-art gait recognition method. We test this on the TUM-IITKGP gait recognition database and show a significant performance gain in the case of occlusions.

1 INTRODUCTION

Person identification using gait information has become an established field of research. In order to identify people, current applications successfully use physiologic features such as face, iris and fingerprint. However it is also possible to detect people using behavior based features such as voice, dialect, signature and gait. The main advantage of using gait features over other features (like face, iris, fingerprint) is the possibility to identify people from large distances and without the person's direct cooperation. For example, in low resolution images, a person's gait signature can be extracted, while the face is not even visible. Also no direct interaction with a sensing device is necessary, which allows for undisclosed identification. Thus gait recognition has great potential in video surveillance, tracking and monitoring.

A major challenge for recognition of people using gait is that almost all current approaches (Han and Bhanu, 2006)(Lee and Grimson, 2001)(Wang et al., 2003) need a sequence of the video, where a complete gait cycle of the walking person is visible. A complete gait cycle is a sequence starting with one foot forward and ending with the same foot forward.

In most databases this is not a problem, because the databases are constructed to always show the complete gait cycles. However in actual real world applications, fully visible gait cycles can not always be guaranteed. Assume for example an airport, a train station or other crowded places. In these scenarios, the gait cycles can be corrupted due to occlusions by other walking pedestrians. Current approaches fail in these cases.

We thus present a method to overcome these limitations. Therefore we propose a preprocessing stage, which effectively performs motion segmentation on the input video and reconstructs synthetic complete gait cycles using partial information available from the corrupted and occluded input sequences.

In principle our methods consists of two separate parts: First, all walking people are detected, tracked and accurately segmented. Thus the first part can be thought of as an application of motion segmentation. Once people are segmented in each frame, in the second part, the gait cycle of each person can be analyzed and complete gait cycles can be reconstructed.

We test our method on the TUM-IITKGP gait recognition database (Hofmann et al., 2011). This database features sequences where each person is completely visible and other sequences where each person is occluded by other pedestrians. On the two baseline algorithms, we show that our preprocessing greatly increases recognition results in the case of occlusions.

In Section 2 and 3 we present the two processing parts. We show results in Section 4 and we conclude in Section 5.

Hofmann M., Wolf D. and Rigoll G..
IDENTIFICATION AND RECONSTRUCTION OF COMPLETE GAIT CYCLES FOR PERSON IDENTIFICATION IN CROWDED SCENES.
DOI: 10.5220/0003329305940597
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2011), pages 594-597
ISBN: 978-989-8425-47-8

2 PERSON TRACKING AND SEGMENTATION

In this section we present our motion segmentation method, which is used to detect, track and accurately segment a varying number of people in the input videos.

We use a graph based image segmentation technique. Here an image is represented as an undirected graph. Nodes correspond to pixels and a cut through connecting edges reveils the segmentation. A similar approach for motion segmentation is for example (Bugeau and Pérez). Our approach, however, is different in that we incorporate an explicit counting of people using mean shift.

We define $\mathcal{P}$ the set of all pixels in a frame and $\mathcal{L}$ the set of possible labels. We denote by $l(p) \in \mathcal{L}$ the labeling for a specific pixel $p \in \mathcal{P}$ and $l(\mathcal{P}) = \{l(p)|p \in \mathcal{P}\}$ the set of all label assignments. $\mathcal{N}$ denotes the standard 4-connected neighborhood and consists of the corresponding pixel pairs (p,q). With these definitions, the Energy definition becomes:

$$\begin{aligned} C(l(\mathcal{P})) &= \sum_{p\in\mathcal{P}} C_{Data}(l(p)) \\ &+ \gamma \cdot \sum_{(p,q)\in\mathcal{N}} C_{Smooth}(l(p),l(q)) \\ &+ \delta \cdot C_{Labels}(l(\mathcal{P})) \end{aligned} \quad (1)$$

The data term describes of assigning labels to individual nodes. The smoothness term ensures smoothness of the resulting objects and the label term discourages the use of too many (small) objects. The data and smoothness term are explained in detail below, the label term $C_{Labels} = |\mathcal{L}'|$ is the number of labels used in the final assignment.

2.1 Data Term

The data term $C_{Data}(l(p))$ describes the costs associated with assigning label $l(p)$ to pixel p. In our work, the data term consists of a term defined by background subtraction and a term defined by optical flow.

$$C_{Data}(l(p)) = \alpha \cdot C_{BS}(l(p)) + C_{OF}(l(p)) \quad (2)$$

For the background model, we use the first frame of the sequence (which is always empty in the used database). We use the YCbCr color space and take the absolute difference in intensity $I(p)$ as the measure. We do efficient shadow suppression by setting $I(p) = 0$ (thus background) for pixels, whose Cr and Cb values are similar to the background model but may have a significant intensity difference. Let l_H be the label of the background and $\mathcal{L}_V$ be the set of all other labels representing people. Then the costs for labeling pixel p with label l becomes

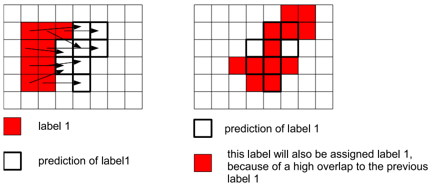

label 1
prediction of label1
prediction of label 1
this label will also be assigned label 1, because of a high overlap to the previous label 1

Figure 1: Predicting label assignment using optical flow.

$$C_{BS}(l(p)) = \begin{cases} I(p) & \text{für } l = l_H \\ (1 - I(p)) & \text{für } l \in \mathcal{L}_V \end{cases} \quad (3)$$

The background term C_{BS} described above is able to separate background from moving objects, however naturally it is not possible to distinguish between multiple walking people. Thus we additionally incorporated an optical flow term C_{OF}, which (1) separates objects moving in different directions and at the same time (2) ensures consistent tracking of objects (i.e. keeping the objects identity)

We first apply mean shift clustering on the output from optical flow (Farnebäck, 2002). This not only finds the number of objects in the frame, but also gives a rough estimate on where the respective objects can be found. More specifically we use a 3-dimensional mean shift, with x, y and ϕ dimensions. Here ϕ corresponds to the direction of the flow at pixel p. We set the size of the used mean shift vector to approximately the expected height and width of a person ($[80 \times 200]$ Pixels), as well as 80° for the flow direction.

Because the mean shift clustering is independent from previous frames, the labels from the mean shift clustering are arbitrary and have to be matched to the labeling of the energy formulation. This is illustrated in Figure 1. First, a predicted labeling is calculated from the previous frame using the optical flow estimate $v(p)$ at each pixel: $\hat{l}(p + v(p)) = l(p)$. Then each mean shift cluster is assigned the label which best fits to the predicted labellings.

Let $\mathcal{P}_l$ be the set of pixels in the mean shift cluster which best fits the original pixels in cluster l. Then the contribution to the data term is defined as

$$C_{OF}(l(p)) = \begin{cases} -\beta & \text{für } p \in \mathcal{P}_l \\ 0 & \text{sonst} \end{cases} \quad (4)$$

Thus, assigning a given label is encouraged by β, if a corresponding object has been found by the mean shift clustering.

2.2 Smoothness Term

The smoothness term models the similarity of the color of adjacent pixels p_1 and p_2. We set it as follows:

$$C_{Smooth}(l(p_1),l(p_2)) = \begin{cases} 0 & l(p_1) = l(p_2) \\ \mathrm{e}^{-\frac{\|\mathbf{c}(p_1)-\mathbf{c}(p_2)\|^2}{\sigma^2}} & l(p_1) \neq l(p_2) \end{cases} \tag{5}$$

Here, $\mathbf{c}(p_i)$ is the color vector of pixel p_i.

2.3 Energy Minimization

For minimizing the energy of 1, we use the α-expansion algorithm (Boykov et al., 2001)(Kolmogorov and Zabih, 2004) including label term (Delong et al., 2010). Thus we seek to find the optimal labeling $\hat{l}(\mathcal{P}) = \operatorname{argmin} C(l(\mathcal{P}))$.

3 GAIT RECONSTRUCTION

The second part of the paper describes how to detect occlusions and how to compensate for them. The main idea is to replace silhouettes, which are completely or partly occluded, by silhouettes from other frames, which are not occluded. To this end, two steps are necessary: First all silhouettes which are occluded are found. Then the gait period is calculated and lastly so called ”reconstruction frames” are searched and used to replace the corrupted silhouettes.

3.1 Detection of Occlusions

A good feature to detect occluded frames is the number of pixels in the silhouette. Figure 2 shows the number of pixels in a tracked silhouette. It can be seen that frames between 145 and 150 are definitely occluded. In order to find the precise range of occluded frames we first calculate the median M_P of pixels. Silhouettes which have less than $\frac{1}{2} \cdot M_P$ pixels are definitely occluded. From these silhouettes we search backward and forward until the number of pixels goes above $0.95 \cdot M_P$ and all pixels in this range are also considered part of the occluded range.

3.2 Measuring the Gait Period

For the last step of finding reconstruction frames, it is necessary to have a good estimate of the gait period T. In order to find the gait period, for each frame, the lower half of the silhouette is correlated with the lower parts of the silhouettes in all other frames. The difference in frame number to the frame with the best match is recorded in a list. The median frame difference in this list yields the gait period we are looking for. This relatively simple method performs very reliably and correctly finds the gait period for all the sequences in the database.

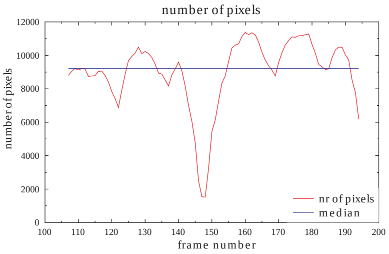


Figure 2: Number of pixels in a silhouette.

3.3 Finding Reconstruction Frames

The goal is to reconstruct the silhouettes which are corrupted by occlusions. The procedure is best described using Figure 3. Here, frames 48-50 (red) are the occluded frames which are to be reconstructed. We seek to find silhouettes from other frames which are similar to the occluded ones. This is possible, because it can be assumed that the sequence consists of multiple gait cycles, which are very similar to each other. Thus the corrupted frames can be replaced by corresponding frames from another gait cycle. Of course, the corrupted frames themselves cannot be used to find corresponding frames. Thus we take the last non-corrupted frame (frame 47 in Figure 3) and search for a best match using normalized correlation. It is important to note that this search may not be performed on all other available frames, but only on those which are in a search region, either one or more gait periods ahead or in the past. This is necessary to avoid incorrect matches for example to a silhouette where the opposite foot is extended. Thus, the possible search ranges are $S_k = [kT - \Delta, kT + \Delta], \quad \forall k \neq 0$. Once the best matching frame is found, the corresponding frames are copied to replace the corrupted frames.

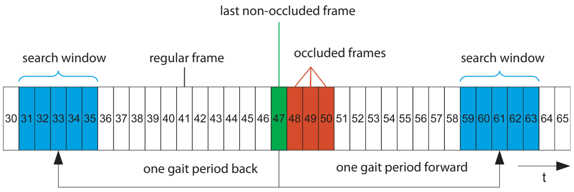


Figure 3: Search area for finding the reconstruction frames.

4 RESULTS

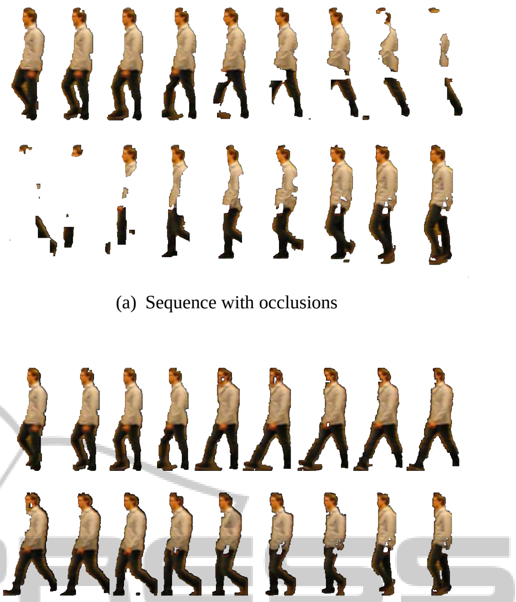
To our knowledge, not much work has focused on gait recognition in the presence of occlusions. However it is a very important issue when moving towards a real working gait recognition system. There are many gait recognition databases, e.g. HumanID, UCSD, CMU Mobo, Soton, CASIA, and others (Hofmann et al., 2011). We use the TUM-IITKGP gait recognition database (Hofmann et al., 2011), because it focuses on gait recognition with occlusions. This database features 35 individuals which are recorded in six different configurations (Regular, hands-in-pocket, backpack, gown, dynamic occlusions, static occlusions). In this paper we use the first configuration for training and focus on the last two configurations for recognition. For all our experiments we set $[\alpha = 200, \beta = 80, \gamma = 30, \delta = 40.000]$.

We compare our results to the two baseline algorithms described in (Hofmann et al., 2011). In Table 1 evaluation results are shown. The first algorithm (based on color histograms) is clearly outperformed. The second baseline algorithm (based on Gait Energy Image) could not work without the preprocessing we propose in this paper.

Table 1: Top 1 recognition rates for Baseline 1 (Color Histogram) and Baseline 2 (Gait Energy Image).

	dynamic occlusions	static occlusions
Baseline 1	43.7%	70.0%
Baseline 1 + ours	84.3%	87.5%
Baseline 2	-	-
Baseline 2 + ours	67.2%	72.7%

Qualitative results of our proposed reconstruction method are shown in Figure 4. It can be seen that the corrupted sequence is nicely reconstructed.



(a) Sequence with occlusions

(b) Reconstructed sequence

Figure 4: Example of reconstruction (b) of gait cycle which was originally (a) corrupted by occlusions.

5 CONCLUSIONS

In this paper we have shown a preprocessing stage that allows to reconstruct complete gait cycles such that gait recognition is possible in spite of occlusions. In principle this preprocessing can be applied to any gait recognition algorithm. In our experiments we have shown that a preprocessing like ours is in fact beneficial in the case of occlusions.

REFERENCES

Boykov, Y., Veksler, O., and Zabih, R. (2001). Efficient approximate energy minimization via graph cuts. *IEEE TPAMI, Number 20(12):1222-1239.*

Bugeau, A. and Pérez, P. Track and cut: Simultaneous tracking and segmentation of multiple objects with graph cuts.

Delong, A., Osokin, A., Isack, H. N., and Boykov, Y. (2010). Fast approximate energy minimization with label costs. CVPR.

Farnebäck, G. (2002). *Polynomial Expansion for Orientation and Motion Estimation.* PhD thesis, Linköping University, Sweden.

Han, J. and Bhanu, B. (2006). Individual recognition using gait energy image. *IEEE TPAMI, Volume 28, Number 2.*

Hofmann, M., Sural, S., and Rigoll, G. (2011). Gait recognition in the presence of occlusion: A new dataset and baseline algorithms. In *International Conferences on Computer Graphics, Visualization and Computer Vision (WSCG).*

Kolmogorov, V. and Zabih, R. (2004). What energy functions can be minimized via graph cuts? IEEE TPAMI, Number 26(2):147-159.

Lee, L. and Grimson, W. E. L. (2001). Gait analysis for recognition and classification. *MIT Artificial Intelligence Lab, Cambridge.*

Wang, L., Tan, T., Ning, H., and Hu, W. (2003). Silhouette analysis-based gait recognition for human identification. IEEE TPAMI, Vol. 25, No. 10.