Skeleton-Based Segmentation and Recognition of Human Activities from Video Sequences

UNIVERSITÉ LAVAL

precarn

Computer Vision and Systems Laboratory
http://vision.gel.ulaval.ca

Frédéric Jean, Sébastien Quirion, Alexandra Branzan Albu, Robert Bergevin

Abstract / Résumé

This work presents an approach to segment and recognize human activities from skeletons obtained from the binary human silhouettes of a video sequence on a frame-by-frame basis. Temporal variation of the different angles between adjacent parts of the body (such as arm and forearm, or forearm and torso) are used as input signals to the segmentation and recognition algorithms. Segmentation is achieved using a recursive auto-covariance analysis and a periodicity measure on angle signals. For activity recognition, a feature vector is generated with a subset of components of the Fourier transform of angle signals. We then compare activities using a minimum distance classifier and the Euclidian distance.

Le but de ce projet est de segmenter et de reconnaître les activités effectuées par des sujets humains à partir de squelettes obtenus pour chaque trame d'une séquence vidéo. Les variations temporelles de différents angles entre certaines parties du corps sont utilisées comme signaux d'entrées pour l'algorithme de segmentation et de reconnaissance présenté. La segmentation est obtenue à l'aide d'une analyse d'auto-covariance récursive et d'une mesure de périodicité sur les signaux d'angles. La reconnaissance est réalisée en formant des vecteurs de caractéristiques avec un sous-ensemble des composantes de la transformée de Fourier des signaux d'angles. Les activités sont finalement comparées en calculant la distance euclidienne entre les vecteurs de caractéristiques correspondants.

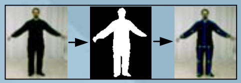
1. Introduction

Intelligent surveillance systems based on computer vision techniques represent an interesting alternative to the traditional CCTV systems. Systems for monitoring indoor and outdoor environments are to perform real-time tracking of human subjects in order to detect and recognize their activities. Complex human activities require a model-based tracking approach. Based on the assumption that the human body in motion has the physical behaviour of an articulated object, our study uses a stick-model (or skeleton). The main objectives of our research are the model-based detection and recognition of periodic human activities.

2. Pre-processing

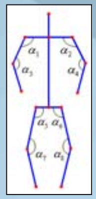
2.1 Skeletonization

A skeleton is computed from a human silhouette for each frame of video sequence with the method proposed in [1].



2.2 Joint Position Filtering

Temporal filtering of skeleton joints is done in order to reduce noise induced by the skeleton fitting process. Temporal filtering is performed for each joint by computing the mean coordinate using the previous, current, and next joint coordinates.



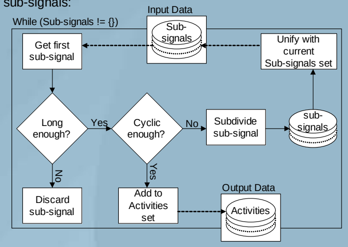
2.3 Angle Computation

Temporal variation of the different angles between adjacent parts of the body are computed (see figure) since it contains relevant information for activity detection, description and recognition.

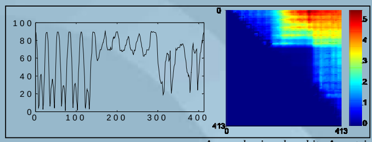
3. Activity Segmentation

3.1 Single signal segmentation

The goal is to iteratively extract activities from a signal using a signal periodicity measure [2] and an auto-covariance analysis. A first step is to suppress any "silences" in the signal. A "silence" is defined as a compact time interval longer than a critical length where the amplitude of the signal does not vary significantly. A set of sub-signals is obtained by suppressing these "silences". A sub-signal is defined as a compact time interval of a signal or of another sub-signal. Activities can then be extracted from this set of sub-signals:



The length test compares the length of the sub-signal to a fixed threshold while the periodicity test compares the result of a periodicity measure [2] on the sub-signal to another fixed threshold. In order to subdivide a signal, its auto-covariance matrix $\boldsymbol{A}$ is first computed. The value of each $\boldsymbol{a_{ij}}$ is the standard deviation of the residuals obtained from a linear regression on the local maximums of the auto-covariance of the sub-signal starting at index $\boldsymbol{i}$ and ending at index $\boldsymbol{j}$. A low value for $\boldsymbol{a_{ij}}$ means that the corresponding sub-signal has good chances of being cyclic. Cyclic portions of the signal in the matrix show as step-like triangles of equally low values on the main diagonal, the tip of each triangle representing the cyclic sub-signal of maximum length. These steps are approximated through binary thresholding. A signal segmentation is then derived and the corresponding sub-signals are generated.

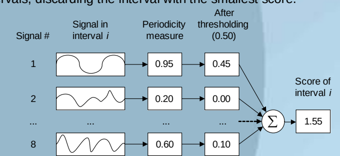


*A sample signal and its **A** matrix*

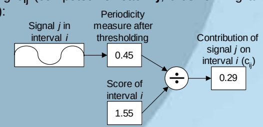
3.2 Signal set segmentation

The objective is to find one optimal segmentation, defined as a set of non-overlapping intervals, that extracts activities from all the signals. Each signal is first segmented in the way described above. The temporal location of each of the activities found in this step are considered as candidate intervals.

This set of candidates has to be modified as it is likely to contain overlapping intervals. A first, simple, method to correct this consists in rating each interval and, for every pair of overlapping intervals, discarding the interval with the smallest score.



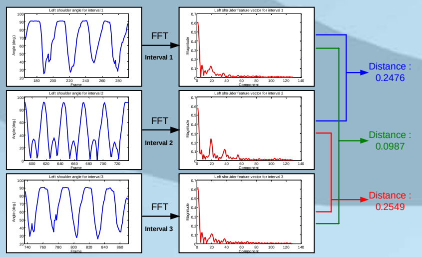
The set of remaining intervals is the final segmentation. A contribution value is then computed for each signal-interval pair generating $\boldsymbol{c_{ij}}$ (computed for each $\boldsymbol{ij}$, thus forming a contribution matrix $\boldsymbol{C}$):



4. Activity Recognition

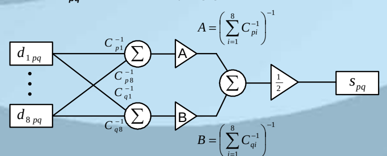
4.1 Describing and Comparing Intervals

Feature vectors are formed for each interval on each angle signal by using the first 25% of the Fourier transform components [3]. Intervals $\boldsymbol{p}$ and $\boldsymbol{q}$ on a signal $\boldsymbol{i}$ are then compared using the Euclidian distance $\boldsymbol{d_{ipq}}$ between their feature vectors.



4.2 Integrating Distances

Distances $\boldsymbol{d_{ipq}}$ are integrated in order to obtain a global similarity measure $\boldsymbol{s_{pq}}$ for each interval pair $\boldsymbol{pq}$:



$$A = \left(\sum_{i=1}^{8} C_{pi}^{-1} \right)^{-1}$$

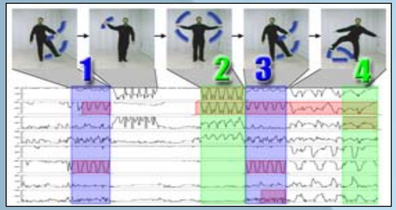
$$B = \left(\sum_{i=1}^{8} C_{qi}^{-1} \right)^{-1}$$

Similar activities will get a low value while different activities will get a high value.

4.3 Thresholding

A threshold is applied to the similarity measures $\boldsymbol{s_{pq}}$ obtained in the previous step : values exceeding a given threshold (see Section 5) are set to 0 (different activities) while the others are set to 1 (similar activities).

5. Results



Example of analysed periodic activities in a video sequence and activity description with angle signals. Candidate intervals are shown in red and final segmentation is shown in blue and green.

The database for this study contains video sequences acquired with a monocular camera, containing periodic activities as shown above.

As shown in the figure above, the final segmentation successfully detects 4 of 5 activities in the sample sequence. The description and recognition step successfully recognize the two instances of the same activity (1 and 3), as shown in the following tables:

Interval	1	2	3	4
1	0	0.2593	0.1504	0.2751
2	0.2593	0	0.2812	0.2838
3	0.1504	0.2812	0	0.3195
4	0.2751	0.2838	0.3195	0

Activity similarity measure ($\boldsymbol{s_{pq}}$)

Interval	1	2	3	4
1	1	0	1	0
2	0	1	0	0
3	1	0	1	0
4	0	0	0	1

Similar activities (threshold 0.2)

6. Future Work

Ongoing work is focused on extending our approach to the detection and recognition of natural quasi-periodic human activities, such as walking and running. To do this, we are currently working on self-occlusion and pose change issues. Automatic thresholds for segmentation and recognition are also part of our current research.

7. Conclusion

This paper presented an approach to model-based segmentation and recognition of periodic human activities. Our iterative approach to activity segmentation, as well as our approach to activity description and comparison, yielded interesting results. While it is clearly too early to generalize our results to a greater set of activities and scenarios, our current results lead us to believe that we can achieve relatively accurate segmentation and recognition even with a sequence of fair complexity and noisy angle signals.

References

[1] J. Vignola, J.-F. Lalonde, and R. Bergevin, Progressive Human Skeleton Fitting. Proceedings of the 16th Conference on Vision Interface, 2003.
[2] R. Polana and R. Nelson : Detecting activities. Journal of Visual communication and Image representation 5(2), 1994.
[3] J. Hayfron et al , Automatic gait recognition by symmetry analysis. Pattern Recognition Letters 24, 2003.

Acknowledgments

This work is supported by Precarn Inc (MONNET Project) and by a FQRNT postgraduate scholarship.

precarn

Fonds de recherche sur la nature et les technologies Québec