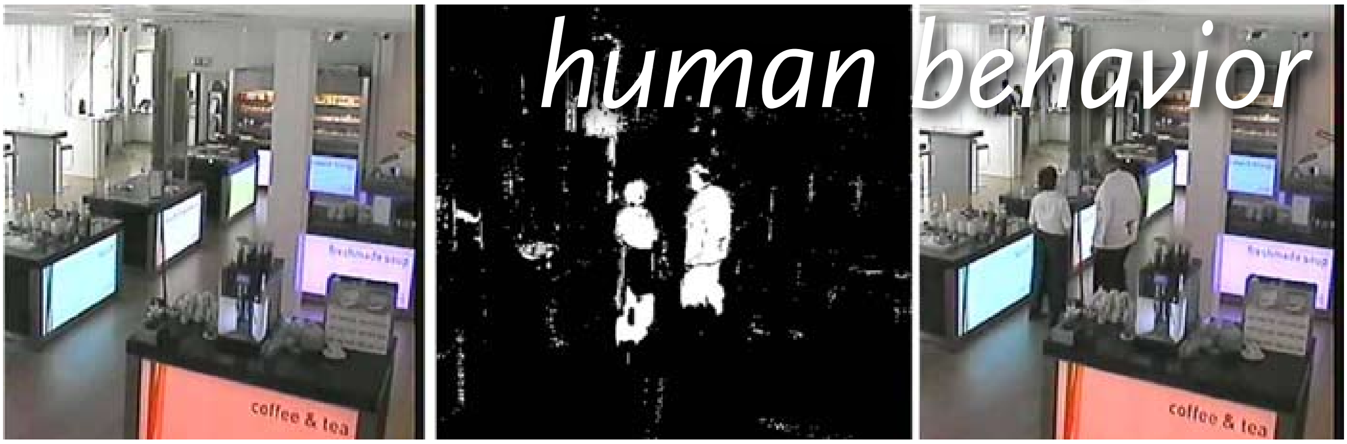
Automatic annotation of human behavior



VidART

A toolbox for automatic human pose and gesture recognition.

Knowledge Transfer Project:
VidART – Video Analysis and Recognition Toolbox

Partners
Utrecht University
Noldus Information Technology BV

Budget
300.000 euro

Contact details
Remco Veltkamp
Utrecht University
Remco.Veltkamp@cs.uu.nl

Human behavior studies are complex and manual annotation of the desired behaviors is tedious. To automate the identification and analysis of human poses and gestures from videos, requirements must be set about the scene. Another prerequisite is a robust method to detect people in the scene.

"Manually human behavior annotation of video is a tedious work"

Noldus Information Technology BV has developed software to manually annotate and analyze human behavior. Automation of human behavior detection can speed up the annotation and analysis process significantly. Using the Restaurant of the Future as a benchmark, requirements are formulated for the system. For any kind of automatic analysis, classification and annotation of the behavior of people from a video, a robust method is needed to detect humans in the scene. For an indoor environment, background subtraction is a well-known detection method, but it faces many challenges such as illumination changes in the scene and shadows casted by moving objects.

Prerequisites of the system

The Restaurant of the Future is taken as a benchmark to list the requirements on the system (hardware, software, environment and objects). For example, the system must be able to deal with varying illumination, both in intensity and color because of the presence of large windows (influence of sunlight), the large amount of light sources and the different colors of these light sources. Another requirement is that persons have no any additional sensors or markers to ensure a natural recording. To detect people, background subtraction uses a model to represent the background scene. Deviations of the current frame with respect to this background model identify the foreground objects. Several background subtraction algorithms have been analyzed on its performance and accuracy. To handle shadows casted by foreground objects and illumination changes in the scene, a multiple view background subtraction method has been investigated, which uses geometrical information about the scene to detect foreground objects. Often background subtraction gives unreliable or incomplete silhouettes. An alternative method based on level-sets has been investigated to find the contour of the object. The first results show that this method is less sensitive to shadows and illumination changes.

From detection to pose and gesture recognition

Ones the detection of persons is done, the persons have to be tracked and identified over time. The main challenge of tracking is (self)-occlusion, which can be handled by using multiple cameras. When the humans are separated, pose estimation can be performed on each person individually. This will result in the position of each body part, like the head, hands, shoulders, legs and torso. The final part of the analysis toolbox will be the recognition of certain pre-defined gestures. The software has to be tested and verified. The easy way is to use publicly available data sets for tracking and pose estimation. However, to have more freedom in our testing (specifically in gesture recognition), a data set will be recorded in the Restaurant of the Future and a ground truth will be provided.