論文の内容の要旨

Abstract

Title of Dissertation:
Human Behavior Analysis for Customer Interest Level Estimation using In-store Cameras
（論文題目：店舗カメラを用いた顧客関心度推定のための人物姿勢検知に関する研究）

氏　　名　　劉　景文
Name of Author　　Jingwen Liu

The customer behavior analysis is one of the most concerned topics for retailers because the customer behavior information can indicate the customer interest level to the product in the stores and is helpful to increase the commercial benefit. Conventionally, retailers can collect the records from credit cards and cash registers to analyze the customers from their buying behaviors. However, these records are not able to reveal the “not buying” behaviors of customers. The customer behaviors inside the retail store are still in the black box. Therefore, this dissertation applies surveillance camera to analyze customer behavior for customer interest level extraction.

In order to assess the customer interest level, this dissertation focuses on 3 main parts: head and body orientation detection, pose estimation and arm action classification. These 3 parts reveal increasing customer interest level. When the customer has an initial interest in some products, he or she probably looks at or turns to merchandise shelf. The head and body orientation can tell these behaviors. If the interest level becomes higher, the customer may bend over or squat down to look at the items carefully. Thus the pose information including the joint positions also needs to be estimated. Additionally, the customer arm actions can show further cues, especially for the interactions between merchandise, such as touching, taking items, returning to shelf or putting into the basket. Therefore, the classification of semantic arm action is one of the essential parts of this system.

In the customer head and body orientation detection part, the orientation is used to indicate

whether the customer is facing or turning to the merchandise shelf. Conventionally, the orientation can be predicted by Supervised Learning (SL) method. However, in the retail store environments, the ground truth of orientation is difficult to obtain and thus the orientation annotations are usually decided subjectively. This results in inaccurate orientation annotations in the training data. In order to solve this problem, this dissertation proposed a Semi-Supervised Learning (SSL) method to automatically reduce the inaccurate annotations. In SSL, the training dataset is annotated as strongly labeled data and weakly labeled data. The training data with ambiguous orientations are given weak labels, while the data with clear orientations are given strong labels. The SSL refines the classifier by grouping the weakly labeled data into strongly labeled data in iterations. This dissertation also incorporates the state-of-art structure ResNet as the classifier. As far as the author, there is no research using deep learning to classify head and body orientation. Additionally, this dissertation also applies the human physical constraints and the temporal constraints in body and head orientation detection in image sequence. The experiments show that the deep learning + SSL methods outperform the conventional feature based methods and feature + SSL methods.

In the pose estimation part, the essential idea for pose estimation is to incorporate the pose with body orientation, joint connections and visibility mask. Firstly, a customer pose estimation network integrating different useful information is proposed, named Integral Pose Network (IntePoseNet). This system firstly generates initial joint heatmaps using Fully Convolutional Networks (FCN). Based on these heatmaps, a series of body orientation based message passing layers construct a Deformable Parts Model (DPM) to model the local joint connections. At the output end, the joint positions and visibility mask are calculated from the joint heatmaps. Secondly, a multi-task neural network is proposed to simultaneously output joint heatmaps, joint connection maps, body orientation and visibility mask. In the multi-task learning, the tasks can improve the performances each other. The experiments show that the multi-task neural network outperforms the conventional DPM method, deep learning + DPM method, the state-of -art deep learning methods and IntePoseNet.

In the arm action classification part, this dissertation defines 3 semantic arm actions: *touching*, *picking and returning* to shelf and *picking and putting* into basket. These actions reveal increasing customer interest in the product. For the arm action classification, this dissertation proposes a novel Combined Hand Feature (CHF), including hand trajectory, hand tracking status and the interactions between hand and shopping basket. The CHF is extracted from every image and then is given input into a Dynamic Bayesian Network (DBN). The DBN is used to classify the CHF sequences into different arm actions. The experiments prove that the CHF outperforms the traditional algorithm using HOG + SVM.