Intelligent Consumer Flow and Experience Analysis System Based on Cognitive Intelligence: Smart Eye System

Chujie Tian, Ruizhi Zhang, Chunhong Zhang and Xiaomeng Zhao
Beijing University of Posts and Telecommunications, Xitucheng Road No.10, Haidian District
Email: ruizhizhang@bupt.edu.cn

Abstract. With the saturation of the market and the competition in e-commerce, the focus of the offline new retail has gradually shifted from commodities to consumers. Although there have been some developments in the offline retail industry in terms of consumer analysis in recent years, such as consumer flow statistics based on the non-visual sensors and consumer experience based on the questionnaire, it is still difficult to meet the growing demand due to the limitations of traditional technologies. To deal with the problem, Smart Eye System (SES) is proposed to analyze the characteristics and evaluate the experience of consumers by Cognitive Intelligence (CI). As a new attempt, face recognition, emotion recognition and ambient information are combined to provide more detailed and accurate analysis of consumers, including the statistics, characteristics and experience of consumer flow. As the first batch of products that combine CI with commercial applications, SES brings not only more accurate analysis and broader perspectives to the offline retail industry, but also a frontier experiment that commercializes CI.

1. Introduction
Due to the development of the market and e-commerce, more varied options have been brought to the consumers. The transition from commodity-centric to consumer-centric has been regarded as one of the most important changes to maintain the vibrancy of the market [1]. Analysing the characteristics and experience of the consumers is urgent.

As an important indicator which can influence business decisions, the statistics of consumer flow is deeply concerned [2]. It evolved from early artificial statistics to automatic machine identification [3]. Infrared sensors, which are common devices for identifying the number of consumers, are widely deployed in malls. In addition, to understand the specific composition and satisfaction of consumers, the characteristics and the experience of consumers are more concerned [4]. The acquisition of information is mostly through the manual surveys and the feedback of entity experience store. Obviously, the limitations caused by the existing commercial applications have hindered the development of the offline retail. Using infrared sensors is hard to recognize whether the person has already appeared. The result of consumer experience from questionnaires is not reliable due to subjective feedback from investigators.

Smart Eye System (SES) based on CI is proposed to solve these problems. CI refers to the ability of a system to gradually acquire knowledge through experiences during its existence [5]. SES achieves better analysis of the consumers in the offline retail by face recognition and emotion recognition. Consumers who occur repeatedly can be accurately identified by face recognition. Identifying and analyzing the age and gender of consumers in real time helps the offline retail make business decisions. Emotion recognition is used to analyze the experience of consumers objectively and continuously.
The reminder of article is organized as follows: Section 2 reviews related work in SES. Section 3 presents the features and principles of SES. Section 4 describes framework and components of SES. Section 5 shows specific cases in the department store. Section 6 concludes this system and outlines further work.

2. Related Work

2.1. Consumer Flow Statistics
The earliest statistical method of consumer flow is counting by staff, the disadvantage of which is the inability of employees to work long-term. To meet the demand for continuous and accuracy data, automatic consumer flow statistics system which includes three roll gate mode, gravity induction mode and infrared induction mode appeared. With the improvement of computer image processing capabilities, consumer flow statistics technology based on image is researched in full swing. The methods can be divided into two categories: the indirect and direct method. A function is constructed to map between the scene feature and the number of people in the indirect method [7]. Various features are extracted from each foreground segment together with the corresponding key points which are highly correlated with the size, density, and occlusion level of the crowd [8]. The direct method is to detect the position of each face directly, and then track the face [9]. Due to the high cost of tracking faces [10], body detection is mostly applied by vertical shooting in real scenario, which ignores characteristics and experience of consumer.

2.2. Consumer Experience
Consumer experience data comes generally from reviews online, questionnaires and feedback from entity experience stores. The former two cannot guarantee the reliability of the data. Opening entity experience stores costs human and material resources, which is not applicable to all products. In addition, there is no scientific and unified method to cover every consumer in experience analysis. Therefore, the traditional consumer experience analysis methods have obvious defects in data reliability and scientific reference.

2.3. Cognitive Intelligence
Cognitive intelligence (CI) is a kind of technology which is based on the ability of machines to sense, reason, and act and adapt based on learned experience, as Intel CEO Brain Krzanich said. After decades of development, CI has improved dramatically with impressive gains in computer vision, natural language processing, speech recognition, and robotics, among other areas. As one of the technologies, face recognition was originally proposed to use in security systems. From an image of the subject's face, facial features include the relative position, size, shape of the eyes, nose, cheekbones, and jaw are extracted [11]. With the development of machine learning, neural networks and deep learning, the accuracy of face recognition has improved a lot [12], which makes it feasible to apply in some fields such as face payment. Emotion recognition is to recognize expression from face-recognized image. The most common approach is to classify continuous expressive facial displays according to specific labels, categories or dimensions. According to Ekman and his colleagues [13], six basic facial expressions of emotion can be received universally: happiness, sadness, surprise, fear, anger and disgust. SES is the first application in retail industry to analysis consumer experience by using emotion recognition technology.

3. Features and Principles of SES

3.1. The Features of SES
The purpose of intelligent consumer analysis system is to identify and analyse consumer information based on CI. Identifying consumer, recognizing consumer characteristics and analysing consumer satisfaction is indispensable for management in new retail. In order to meet these demands, face detection, characteristics recognition and emotion recognition are adopted in SES.
3.1.1 Face Detection. Accurate counting caused by face detection is one of the features in SES. Unlike traditional technology that counts total times consumers pass by, SES can count the real number by identifying consumers through face detection. Besides, the staff in the scenario will not be counted. The time and times each consumer visits can be recorded by SES.

Based on the accurate counting, KPI overview and alarm services are provided in SES. The KPI overview shows the current consumer flow indicators. It also compares with the previous data to analyse the short-term fluctuations and the trend of consumer flow change at the designated location. The purpose of the alarm service is to remind the mall management that the current consumer flow is too small. When the consumer flow is lower than the expected threshold during a certain period, a message will be pushed to warn the user, which includes the location, the current period, and the text description such as the weather and whether the day is a holiday.

3.1.2 Characteristics Recognition. Rich dimension of consumer characteristics is the second feature of SES. Based on CI, gender and age of consumer can be identified by characteristics recognition. Effective analysis of consumer characteristics with ambient information in the dimension of time and space can be provided in SES.

Based on consumer characteristics, group analysis and member arrival reminding service are provided by SES. Group analysis is to statistically analyze the distribution and preference of each group. The number of male/female consumers, household consumers and consumers at different age groups such as the number of young women in a specified area over a certain period of time are calculated. The member arrival reminding is dedicated to provide more detailed services for senior members. Based on the facial characteristics of the members, SES detects their entry into the store and reminds the staff to receive them.

3.1.3 Emotion Recognition. Continuous and real-time consumer experience obtained by emotion recognition is another feature of SES. One of the major advantages of the approach is that objectivity can be effectively ensured through machine identification. Another advantage is that emotion recognition can be obtained and analyzed without disturbing consumers.

Based on emotion information in the results of cognitive analysis, experience can be obtained with the ambient information. The analysis includes statistics and comparison of consumer experience, such as the statistics of overall consumer satisfaction, the observation of specific unsatisfied consumers, and the rank of areas with higher satisfaction or unsatisfactory experience. It is helpful for managers of the offline retail to have a clearly understanding of their own level of services, and quickly solve the possible problem.

3.2. Principles of SES

Principles of SES are shown in Figure 1. First, devices are used to capture consumer facial images and the scenario data are collected from devices and text. Second, the image data are transformed into emotion information, face information and characteristics information by CI. At the same time, promotion information, weather information and holiday information are extracted from scenario data by Natural Language Processing. Third, the above information is combined to count consumer flow, identify consumer characteristics and analyze consumer experience. Finally, the results will be clearly presented by the reports.
4. Framework and Components of SES
The overall framework of the SES is shown in Figure 2. The system consists of data acquisition, data transmission, data procession and cognitive analysis. The function of these modules is as follows:

Figure 1. Principles of SES

Figure 2. Overall framework of SES
4.1. Data Acquisition.
Hikvision smart camera is used in SES to collect data, whose highest resolution can up to 2 million pixels (1920 x 1080), and 30fps real-time image is output at that resolution. It also supports the function of fogging, electronic anti tremor, which is a good access for image and video resource. Because of the low demand for low light scene in retail scenes, the ability of infrared photography is not important. But the camera is very sensitive to reverse sunlight scenes, so it should be avoided being deployed in the reverse sunlight scenes as far as possible. We have refitted the Hikvision camera's SDK to enable it to photograph, video or record at any time user needs.

4.2. Data Transmission and Data Procession
FFMPEG is used to process the video stream on the server side in the mall whose network environment can support RTSP streaming media data transmission. When using FFMPEG for image acquisition, a multi-threaded processing scheme is adopted to open up a separate thread for each image, which solves the problem of time delay when large numbers of cameras working at the same time and accelerates the processing rate. After getting the image resource, the images are uploaded to Azure Block Storage and the information to the Azure database. A buffer is built based on Redis between the acquisition and the upload. SES applies the producer-consumer model to decouple the two models and process them asynchronously to enable them to support larger concurrency.

When the mall network environment does not support RTSP streaming media data transmission, image transfer script is deployed in the mall to process the video stream, whose processing logic is similar to the above, except that the SDK used in the image acquisition module is refitted. In this way, the script periodically requests the server for the screenshot rule, invokes the corresponding executable file for image acquisition according to the rule, and then uploads the corresponding information to the cloud.

4.3. Cognitive Analysis.
The framework of the cognitive analysis module is shown in Figure 3. Images are detected to identify whether the face exists. The faces will be queried if it has already appeared in the same day, which is called uniqueness detection. The face information represents its identity and is unique. Emotion information includes eight categories: "anger", "contempt", "disgust", "fear", "happiness", "neutral", "sadness", "surprise".

![Figure 3. Cognitive analysis of SES](image)

4.3.1 Face Detection. Face detection refers to the process of determining the position of all existing faces in the image. In 2001, Paul Viola and Michael Jones proposed a face detection algorithm based on Adaboost to ensure high processing rate and good detection results [14]. Adaboost is a learning algorithm that builds an accurate classifier. It shifted a set of weak learning algorithms into a strong learning algorithm through certain rules, thus can get a classifier that can achieve a better recognition accuracy through sample training.
4.3.2 Uniqueness test. Uniqueness detection is the module to analyze and identify the similarity faces according to the images in the database. Five steps are designed:

- Facial feature location.
- Face geometric normalization
- Face illumination normalization
- Feature extraction
- Similarity measure

The system can automatically locate the key feature points of face by AdaBoost algorithm. According to the facial feature, face area image is obtained to make it corresponding to the pixels of different faces. To make the images of the same person taken under different conditions (light intensity, direction, distance, posture, etc.) consistent, gamma correction is adopted. The Euclidean distance \((d_{x,y})\) is calculated as follows by the features extracted from the faces, and threshold is set to be the measure of the similarity.

\[
dist(X, Y) = \left( \sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}
\]  

4.3.3 Face Recognition. The convolutional neural network (CNN) is a multi-layered neural network and usually a part of it is a convolutional layer and another part of it is a fully connected layer [15]. CNN has been proved to be effective in solving image classification tasks. For example, it shows the most advanced performance on the ImageNet Challenge. Face recognition can also be seen as an image classification task. The number of classifications it requires is much smaller than the Image Net Challenge [16] (gender: 2 categories, emotion: 8 categories, age: 8 categories vs Image Net: 1000 categories).

A suitable CNN network architecture is proposed to make good generalization capabilities for feature extraction of age, gender, and expression [17]. The network structure is shown in Figure 4. The input image is uniformly trimmed to 227*227. The trimmed image passes through three convolution layers. Each convolution layer also contains pooling and non-linearization. In addition, two full-connected layers are used to map the learned "distributed feature representation" to the sample tag space.

Figure 4. CNN network structure model

5. Case Study

SES is used at Daxing Wangfujing Department Store in Beijing at the end of October 2017. A total of 21 face cameras were installed to cover all important business areas such as entrances, escalators and key service counters. Some positions are shown in Figure 5. The cameras’ angle and position are adjusted according to the quality of the picture, which is helpful to make the faces unblocked. The results are shown in Figure 6.
After the installation and adjustment of the camera, CI is used to analyze the images taken. Due to trade secrets, the interfaces shown below are processed data.

5.1. An Example of Consumer flow
As shown in Figure 7, the consumer flow is presented in the form of the visual report. By comparing the value of the current consumer flow over a period time, the trend has been clearly shown.

Figure 5. Devices installation

Figure 6. The pictures taken by the camera

Figure 7. Consumer flow overall
5.2. An Example of Consumer Flow Feature Recognition

The page shown in Figure 8 presents the different analysis result from the characteristics of the consumer with the space information. As seen in the picture, the first part shows the floors with the largest number of female/male/family consumers in the month; the second part shows the distribution of consumer flow age in different floors, and details are shown in Figure 9. The feature ratio is shown by the terms of circular chart in different floors, and the composition of the consumer group can be clearly identified; the third part shows the proportion of female/male/family consumers in five floors by the terms of pentagon graph, the attribute preference can be found in each group through comparison.

![Figure 8. Floor Analysis](image1.png)

![Figure 9. The distribution of consumer flow of different ages](image2.png)

5.3. An Example of User Experience Analysis

Low-satisfaction-consumer information is displayed in Figure 10, which includes images, consumer gender, age, and location, time, and dissatisfaction levels. Through finding the lowest level of consumer satisfaction among different floors, the consumer dissatisfaction information will be sent to the manager in real time. Figure 11 shows the real-time update of relevant data based on the upper left corner of the video: As the number of people in the video continues to increase, the number of consumers will increase, and the proportions of consumers’ sex, age, and satisfaction will also change concurrently.
Figure 10. Low satisfaction consumer warning

Figure 11. Real-time detection of the satisfaction of consumer flow

Figure 12 shows one specific way to analyze the consumer experience in SES. While the information and location of dissatisfied consumers are identified, the information will be sent to the manager staff to provide the feasibility of quickly finding and solving problems.

Figure 12. Specific cases of consumer experience
6. Conclusion
As the first attempt to combine CI with consumer flow analysis, SES has already succeeded in commercializing in the offline retail. Face detection, characteristics recognition and emotion recognition are combined with ambient scenario information to provide more detailed and accurate analysis of consumers by SES. As the first application to commercialize CI, SES makes it more quick and clear to locate and solve problems. There is still tremendous imagination for offline retail development in combing CI with consumption information.

7. References
[1] Bogers M, Hadar R, Bilberg A, et al. Additive manufacturing for consumer-centric business models: Implications for supply chains in consumer goods manufacturing [J]. Technological Forecasting and Social Change, 2016: 225-239.
[2] Perdikaki O, Kesavan S, Swaminathan J M, et al. Effect of Traffic on Sales and Conversion Rates of Retail Stores[J]. Manufacturing & Service Operations Management, 2012, 14 (1): 145-162.
[3] Schneider R J, Arnold L S, Ragland D R. Methodology for Counting Pedestrians at Intersections [J]. Transportation Research Record: Journal of the Transportation Research Board, 2009, 2140(1): 1-12.
[4] Niu W. Influence Factors of Consumer Satisfaction on Online Consumer Behavior [J]. Value Engineering, 2012.
[5] BOGDAN RADUCANU. FACE RECOGNITION BY ARTIFICIAL VISION SYSTEMS: A COGNITIVE PERSPECTIVE [J]. International Journal of Pattern Recognition & Artificial Intelligence, 2008, 22 (05): -. 
[6] Yahiaoui T, Khoudour L, Meurie C. Real-time passenger counting in buses using dense stereovision [J]. Journal of Electronic Imaging, 2010, 19(3): 031202-031202-11.
[7] Garcia-Bunster G, Torres-Torriti M, Oberli C. Crowded pedestrian counting at bus stops from perspective transformations of foreground areas [J]. Computer Vision, IET, 2012, 6(4): 296-305.
[8] HASHEMZADEH M, FARAZADEH N. Combining keypoint based and segment-based features for counting people in crowded scenes [J]. Information Sciences an International Journal, 2016, 345[C]: 199-216.
[9] Chan A B, Vasconcelos N. Counting people with low-level features and Bayesian regression[J]. Image Processing, IEEE Transactions on, 2012, 21(4): 2160-2177.
[10] Yam K, Siu W, Chan C, et al. Effective bi-directional people flow counting for real time surveillance system [J]. IEEE Transactions on Consumer Electronics, 2011: 863-864.
[11] Peter L. Automated facial coding software outperforms people in recognizing neutral faces as neutral from standardized datasets [J]. Frontiers in Psychology, 2015, 6(1386): 1386.
[12] Zhu Z, Luo P, Wang X, et al. Deep Learning Multi-View Representation for Face Recognition [J]. arXiv: Computer Vision and Pattern Recognition, 2014.
[13] M. Pantic, L. J. M. Rothkrantz, Automatic analysis of facial expressions: The state of the art, IEEE Tran. On Pattern Analysis and Machine Intelligence 22 (12) (2000) 1424–1445.
[14] P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features. IEEE Conf. Computer Vision and Pattern Recognition, Kauai Hawaii, USA, 2001.
[15] Li, G.; Hoi, S. C.; Chang, K.; and Jain, R. 2010. Micro blogging sentiment detection by collaborative online learning. In ICDM, 893–898. IEEE.
[16] Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In NIPS, 4.Age and Gender Classification using Convolutional Neural Networks
[17] H. Han, C. Otto, and A. K. Jain. Age estimation from face images: Human vs. machine performance. In Biometrics (ICB), 2013 International Conference on. IEEE, 2013. 2