Improving retail efficiency through sensing technologies: A survey☆

M. Quintana, J.M. Menendez, F. Alvarez, J.P. Lopez

ABSTRACT

Measuring customer reaction to new products for understanding their level of engagement is necessary for the future of retail. This work introduces a workflow to improve the quality and efficiency of the retail establishments in order to increase their attractiveness. Different research fields are explored to fulfill the requirements of the proposed scenario, remarking most useful works in the literature. The manuscript goes through relevant algorithms to infer properties of consumers like demography, attention or behavior based on their appearance (computer vision techniques), and on signal captured by generation sensors from smart mobile devices, that will add a great value for the retailers. This paper offers a complete overview of this research field which covers a great variety of innovative tools.

1. Introduction

The traditional brick-and-mortar shops are actively looking for completely new ways to attract customers. Smart retail is a term used to describe a set of smart technologies which are designed to give the consumer a greater, faster, safer and smarter experience when shopping. To achieve that pleasing experience, valuable insights should be provided to help clients meet their goals by introducing new ways to understand movement of customers, intentions and their shopping profile, and further more, to offer more personalized services. The trend is exponentially increasing and brands, network aggregators, and needs of media planners are moving toward inferring the level of engagement of customers in order to attract them to new products that satisfy their needs. Future stores should offer what is relevant for the customers at the time they receive it.

Automatic visual data analysis is a very challenging problem. In order to detect humans in video streams and automatically infer their features, interactions or intentions based on their behavior, different computer vision algorithms are required, very often combined with machine learning techniques. The first goal of video analytics in the proposed area of research is to track all the movements of the people that get inside one store. Correlation between time and space for all of them on their shopping experience should be precisely stored by the system. That information can be estimated by standard cameras and image processing algorithms, but the appearance of new kind of sensors like iBeacons or Radio Frequency IDentification (RFIDs) is leading to new hybrid systems with a great performance. Location of potential costumers allows the system to establish a higher abstraction level and provide deeper assessments related to the impact, emotions or synergies incited by the properties of the store and the products exhibited for the shoppers. Basic features extracted from their appearance (like gender, age or ethnicity) should also be determined by the method developed in order to enclose target audience and match its preferences.

The targets specified above require retailers to be proactive in managing and utilizing corporate data if they want to keep up with. There are different motivations to consider business intelligence relevant for smart retail:

- Identifying relationships among the information, and learn how different factors affect each other and the bottom line of the company.
- Companies need to simultaneously analyze multiple layers of information to better understand customer needs and behaviors, coming from analysis down to the individual point-of-sale record.
- A retailer will have many people in different locations with distinctive skills who need to use this information for varying purposes.

IBM developed an interesting overview of the topic in [29] where they propose three main parameters to increase the impact of business analytics: people counting, hot zone and dwell time analysis and customer behavior analysis.

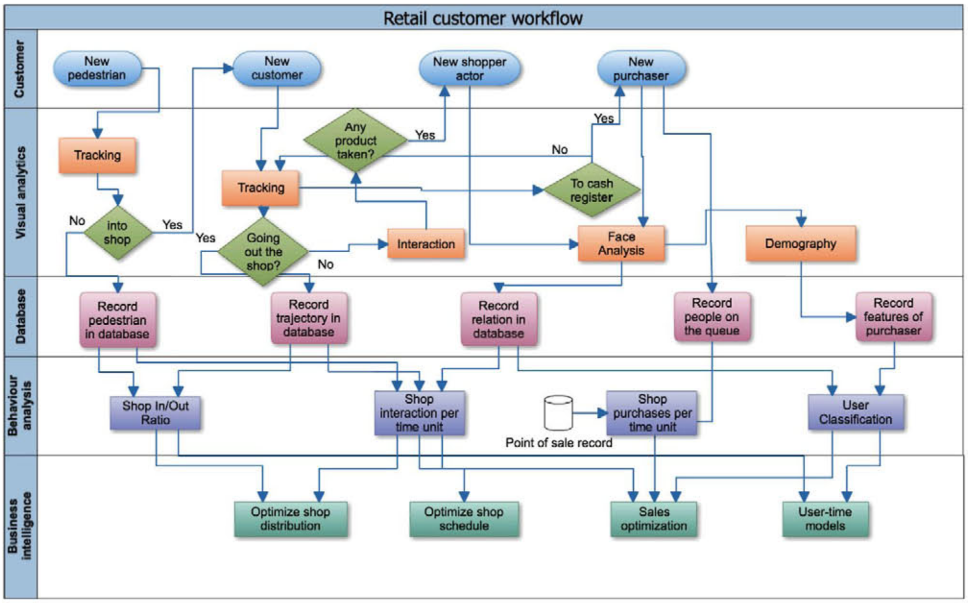

Fig. 1. Workflow to express system response for actions of the consumers.

2. Workflow proposed

This section exposes the workflow that satisfy the needs of the retailers to implement a smart retail in our days. The response of the system for the basic scenario of one shopper walking inside the store can be noticed in Fig. 1. There are five layers designed to provide the different functionalities required by the system:

1. Consumers, there are four states associated to this layer: pedestrian, consumer, shopper actor and purchaser. They are ordered according to their immersion in the retail store.
2. Video analytics, four modules are included in this layer: tracking, interaction, face detection and demography. Tracking module is the preliminary step for the other modules because they rely on the position of the people inside the shop. This module should answer the next questions: is the pedestrian getting into the shop?, are the consumers going out of the shop? and are the consumers going to the Point Of Sale (POS)? The goal of the interaction module is the detection of any relation between the shoppers and any of the products exposed. Finally the face detection step is the preprocessing step of the demography module that extract the features from the purchasers. The source camera for those two modules is usually located close to the POS, and it is also used to optimize the queue operating mode.
3. Database, the objective of this layer is the persistence of the system. Five records will be maintained to keep data ordered and updated: pedestrians, trajectories, relations, people on the queue and features of the purchaser. Another record that stores sales of the store is located outside of the optimization workflow proposed because of its criticism.
4. Behavior analysis, different ratios will be calculated by the system in order to synthesize the behavior of the consumers: In/Out, Interaction/time unit, Purchases/time, shopper attraction unit and shopper classification.
5. Business intelligence, last layer will take the data inferred from the previous layers to predict the improvements that could be applied to increase the efficiency of the establishment. The basic fields to be improved are: distribution, schedule, sales optimization and user-time models.

Different experiments have been performed to analyze intentions of the potential purchasers and/or to attract more consumers:

- Badrinarayana et al. [80] relies on congruity between the multichannel retailers land-based and online stores, obtaining strong implications within the data collected.
- Papagiannidis et al. [93] and Sylvain et al. [63] try to infer from shoppers their purchase intentions, the first of them by introducing simulated products, and the second one by analyzing behavior of the user.
- Finally another methods like [104] or [42] based on incentivation, propose submission of coupons in almost real time through smart mobile devices, and day-ahead dynamic prices to attract more consumers respectively.

3. Survey

This section will go through all the recent and relevant works related to the concept of smart retail shops divided in the layers that have been previously revealed (with the exception of databases and business intelligence since they are out of the scope of this publication)

3.1. Features of the consumers

The main target of one retail establishment is to attract the attention of the consumers and to satisfy them with the provided products and services. All the solutions explored to measure satisfaction of the users implement a preprocessing step focused on the localization and normalization of the Region Of Interest (ROI), where the object of analysis (face of the subject in our case) is mentioned. Khryashchev et al. [10] propose an interesting framework for video analysis with audience measurement purposes. In the explored literature, head pose estimation and gaze analysis are another relevant fields of study not included in the block diagram that will be explored in this manuscript.

3.1.1. Face detection

A good introduction to face detection is included in [61]. It enumerates the parameters with a high influence on the variation in the appearance of human faces that should be considered in this

Table 1
Res. of facial alignment techniques from [8].

Alignment technique	10 fold (%)	Accuracy (%)
Face detector	76.24	68.92
Two points based	79.29	77.86
Three or more points	89.10	82.51
ASM	88.89	84.36
Deep funneled	89.48	80.94

area of computer vision: image orientation, facial expression, head pose, occlusion and illumination. It also presents the different kind of methods that have been applied to solve face detection: feature-based, appearance-based, knowledge-based, and template matching method. In addition, it presents the most widely used solution nowadays: haar feature-based cascade classifier [91]. This method uses integral images to increase the efficiency of the system, and it is supported by a cascade machine learning algorithm that increases its efficacy.

$$feature_i = \sum_{i=1}^{N} w_i RecSum(x_i, y_i, w_i, h_i) \tag{1}$$

where $RecSum(x_i, y_i, w_i, h_i)$ is the summation of intensity in any given upright or rotated rectangle enclosed in a detection window and x_i, y_i, w_i and h_i are for coordinates, dimensions, and rotation of that rectangle.

Güven et al. also provide an innovative application in this field for audience measurement [9]. What they assume is that the face occurrences are limited in this scope and one classifier is able to capture all of them by the combination of Multi-scale Block Local Binary Patterns [13] and Gentle Boost for statistical estimation. Their main contribution is in terms of efficiency to achieve almost real time performance without losing a significant rate of efficacy. Farinella et al. propose a re-identification [4] in order to offer more personalized contents to audience in an advertising environment, based on the prior knowledge of their previous visits. For that aim they build face representations spatial-based on distributions of Local Ternary Patterns (LTP) [99], and claim that a set of M = 50 (training images of the person) and N = 6 (testing face representations) is enough to get satisfactory results.

3.1.2. Image orientation

Once the ROI is located, the cropped image should be normalized to provide relevant information for the next steps. A comparison among the most recently procedures applied for face alignment in a video analytics application is shown in [8]. Three solutions for feature points selection are conferred: two points (eyes), three or more points (eyes corners, nose tip and mouth corners) and Active Shape Model Alignment (ASM) [84]. An open source application (STASM [53]) is proposed to build the model based on Delaunay triangulation [1]. Face alignment algorithms are tested for gender classification with two machine learning techniques: Random Forests (RFs) [2] and Adaboost [18]. The best results are obtained for the combination of ASMs and the second of them, and the poorest performance is achieved by the two points feature points selector. Results for Adaboost classifier on publicly available Labeled Faces in the Wild (LFW) database [76] can be seen in Table 1.

3.1.3. Facial expression

Facial expression is one of the main clues to determine the satisfaction of the consumers. First solution explored in this field [45] relies on the information extracted from the 22 points provided by the Nevenvision facial feature tracker (licensed from Google, Inc.) to classify facial expressions based on five Action Units (AUs) defined by the combination of the codes defined in The Facial Action Coding System (FACS) to represent the contraction of a specific set of facial muscles [32]. Depending on the targeted AU, the proposed algorithm uses HOG features [30] as input for machine learning techniques like RFs, Support Vector Machines (SVM) [22] or Support Vector Regression (SVR) [21] with satisfactory results. In the second one Liu et al. [48] label every image on the training set with 61 points in order to collect frontal visual attributes of subjects. They are used as evidences for the Bayesian Networks (BNs [6]) that performs multi-emotion tagging on video based on a previous multi-expression detection on images. Given the BNs structure the prior probability and the conditional probability are learned from the training data though the maximum likelihood estimation. After training, the posterior probability $P(Y_i|X = X_1, \dots, X_6)$ of a testing sample is calculated. The set of final emotion tags is composed by a state for every one of them $T = \{t_1, \dots, t_i\}$, and the individual values are calculated in the next manner:

$$t_i = \underset{Y_i}{\text{argmax}}\{P(Y_i = 1|X), P(Y_i = 0|X)\} \tag{2}$$

where $i = \{1, \dots, 6\}$. Tests are completed on MPIIGaze dataset [68] to demonstrate the outperformance of the suggested method. Both solutions presented do not reckon the different possible view angles of human faces captured on a retail (or digital signage) environment like [86] does. A Discriminative shared-space prior is generalized from Gaussian Markov Random Field [20] prior for the single view, proceeding a classification of an observed facial expression. A single manifold X is assumed to be shared among the views, and its shared latent space X is built by minimizing the joint negative log-likelihood penalized with the prior placed over the shared manifold following the next equation:

$$L_s = \sum_{v} L_v log(p(X)) \tag{3}$$

where L_v is the negative log-likelihood of data from view $v = \{1, .., V\}$. The algorithm is validated on both posed and spontaneously displayed facial expressions from three publicly available datasets (MultiPIE [75], labeled face parts in the wild [76], and static facial expressions in the wild [31]).

3D modeling is introduced to determine facial expressions in [103], based on blendshape facial animations introduced by Parke [54]. An intermediate model space is generated, where both the target and source AUs (again based on FACS) have the same mesh topology and vertex number, using landmarks for more accurate and efficient shape fitting. Optimized facial expression model is mapped to the target neutral face. Another interesting approach for video analytics is exposed in [102], introducing a unified probabilistic framework based on Dynamic Bayesian Networks (DBNs) to simultaneously and coherently represent the facial evolvement at different levels. Given the model and the measurements of facial motions, all three levels (feature points around each facial component, AUs induced by the codes expressed in FACS and six prototypical facial expressions: happiness, surprise, sadness, fear, disgust and anger) are recognized through a probabilistic inference. The optimal states are tracked by maximizing this posterior:

$$E_t^*, AU_t^*, X_t^* = \underset{E_t, AU_t, X_t}{argmax}\, P(E_t, AU_t, X_t | MAU_{1:t}, MX_{1:t}) \tag{4}$$

where the E_t node in the top level represents the current expression; AU_t represents a set of AUs; X_t denotes the facial feature points to be tracked; MAU_t and MX_t are the corresponding measurements of AUs and the facial feature points, respectively. Results are obtained against the Cohn–Kanade database [51] and the MMI Initiative facial expression database [74]. The main contribution of this work is for facial feature points and AUs.

3.1.4. Head pose and gaze estimation

An Automatic and robust algorithm for head pose estimation can be applied to many real life applications where human interactions are involved. This method [66] presents the preliminary result of face detection and tracking system, by performing initially one segmentation based on human face skin color intensity, and a contour estimation based on the active snake contour and geometric active contour [78]. Its most relevant contribution is that its results probe an ability to detect and track human faces in several poses in real time with uneven lighting conditions. Zhu and Ramanan [69] describes a unified model for face detection, pose estimation, and landmark estimation in real-world, cluttered images, using a mixture of trees with a shared pool of parts. Every facial landmark is modeled as a part, and global mixtures are used to capture topological changes due to viewpoint. Let us write I for an image, and $l_i = (x_i, y_i)$ for the pixel location of part i. We score a configuration of parts $L = \{l_i : i \in V\}$ as:

$$Shape_m(L) = a_{ij}^m dx^2 + b_{ij}^m dx + c_{ij}^m dy^2 + d_{ij}^m dy \tag{5}$$

$$App_m(I, L) = \sum_{i \in V_m} w_i^m \cdot \phi(I, l_i) \tag{6}$$

where $App_m(I, L)$ and $Shape_m(L)$ are the function to model appearance and shape respectively. w_i^m is a placing template for part i, tuned for mixture m, at location l_i. $\phi(I, l_i)$ is the feature vector extracted from pixel location l_i in image I. dx and dy are the displacement of the ith part relative to the jth part.

Speed performance of the real-time algorithm proposed in [34], where RGB-D devices are employed, had a great impact on the research community. 3D coordinates of the nose tip and the angles of rotation of a range image including a head are estimated using random regression forests. The binary test at a non-leaf node is defined as $t_{f,F_1,F_2,\tau}(I)$:

$$|F_1|^{-1} \sum_{q \in F_1} I^f(q) |F_2|^{-1} \sum_{q \in F_2} I^f(q) > \tau \tag{7}$$

where I^f indicates the feature channel, F_1 and F_2 are two rectangles within the patch boundaries, and τ is a threshold. The real-valued vector formed is composed by an offset vector from the point in the range scan falling on the center of the training patch to the nose position in 3D (X, Y and Z) and the head rotation angles denoting the head orientation (yaw, pitch and roll) $\theta_i = \{\theta_x, \theta_y, \theta_z, \theta_{yaw}, \theta_{pitch}, \theta_{roll}\}$. A large database of 50K, 640x480 range images of faces by rendering a 3D morphable model in many randomly generated poses has been collected in [55]. Quantitative evaluation of this solution on a publicly available database (ETH [33]) achieved state-of-the-art performance. Another relevant approach to infer impact from shoppers is gaze estimation. This field was integrated with head pose estimation in a unified framework by this work [101], where eye location results are improved by a hybrid scheme combining it with head pose estimation. Cylindrical head model introduced by Brown in [24] is used to project reference points extracted at eye location step, and afterwards they are constrained within the visual field of view defined by the head pose. Cazzato et al. [82] made an extensive study on how to introduce this kind of solutions in a unconstrained and noninvasive environment without using any calibration or complex devices. They extended their study on [83], by using 3D face model Candide-3 [77]. They performed 4 experiments on different lighting and content conditions to test their system which takes three parameters to solve gaze estimation: eye center, pupil center and eyeball center, concluding that gaze directions in correspondence of visual stimuli can be considered a characteristic pattern for each human being. Another recent relevant publication in this area [68] presents the MPIIGaze dataset that is significantly more variable than the existing ones with respect to appearance and illumination. A method for gaze estimation using multimodal Convolutional Neural Networks (CNN [11]) has been implemented as well. The authors claim that its work is the first one with a large data set under non-controlled conditions, and the results exposed significantly outperforms state-of-the-art methods, as shown on Fig. 2, where results are also shown for UT Multiview Dataset [62].

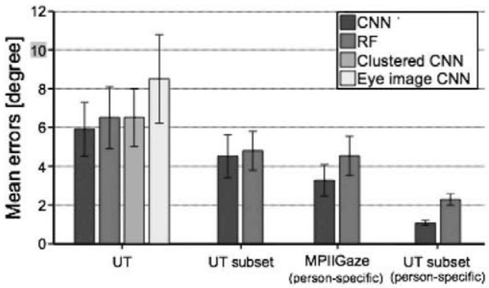

Fig. 2. Results comparison from [68]. Darker color is the intensity chosen for the proposed solution.

3.1.5. Demography

This branch of video analytics is focused on the characterization of the consumers in order to segment them, and to offer them the products that better fit the their preferences. A good and updated survey for gender classification can be found in [60]. It states that considering unconstrained datasets, the feature descriptor better performing is obtained by the combination of HOG, SIFT [50] and Gabor descriptors [58,92]. The results of the approach are tested on unconstrained datasets, like KinFace [73] or LFW [76].

Age estimation is a more complex task than gender estimation, due to the high variability introduced by the aging process. An age estimation system, mostly oriented to audience measurement purposes, is presented in [44]. LBPs are used with an SVM classifier, implementing a multiclass classification approach, and including a new unconstrained face aging database (RUS-FD). In [64] the problem of age estimation is addressed in the context of digital signage. Initial point is classic CLBP descriptor, and it is improved with a bin selection procedure that reduces the computational resources needed, and whose result match the state-of-the-art. Regression techniques are also extensively used for age estimation. Fernández et al. [5] made a rigorous evaluation about the methods belonging to this area of Pattern Recognition: RFs, Multilayer Neural Networks (MNN) [90], Regularized Canonical Correlation Analysis (CCA) [85] and SVR [5]. The accuracy and generalization of each regression technique is evaluated through cross-validation and cross-database validation over two large datasets, MORPH [59] and FRGC [72]. Fig. 3 shows the obtained results. Concluding that the combination of HOG with CCA proved to be the most computationally efficient and straight-forward solution. For applications like retail or digital signage, integration of gender classification, and age and ethnicity estimation in the same solution is very attractive in terms of implementation and computation complexity. In this context, an extensive study on the features that could be applied is included in [3]. Results show that HOG and SWLD [26] descriptors are the most robust and performant among different classification problems. On the innovative solution introduced in [40] that implements the previous statements, Kernel Partial Least Regression (KPLR) is the statistical approach because of three reasons: feature dimensionality reduction and aging function learning are joined, latent variables calculation (real-time applications) and an output vector that contains multiple labels. Kernel variant of the linear Partial Least Square (PLS) [7] model is adopted from [15] (KPLS),

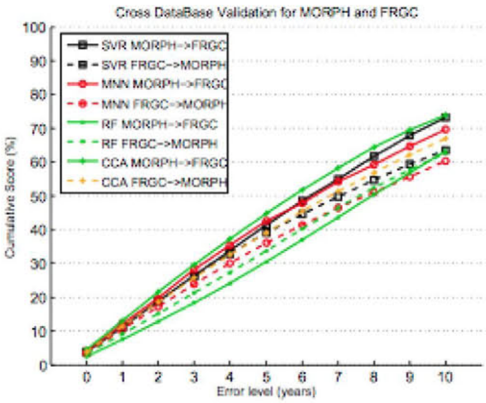

Fig. 3. Results comparison from [5].

Table 2
Res. of age, gender and eth. estimation [38].

Method	Dim	Gender (%)	Ethnics (%)	Age (MAE)
CCA	3	95.2	97.8	5.37
rCCA	3	97.6	98.6	4.42
KCCA	3	98.4	98.9	3.98
PLS	30	97.3	98.6	4.56
KPLS	35	98.3	99.9	4.04

and has the following form:

$$Y = \phi B + F^* \quad B = \phi^T U(T^T KU)^{-1} T^T Y \tag{8}$$

where X and Y are two sets of data with a strong nonlinear relation. K is the Gram matrix of the cross dot products between all mapped input data points, i.e. $K = \phi\phi_t$. ϕ denotes the matrix of the mapped X-space data $\{\phi(x_i) \in F\}_i^n$, where F is the high-dimensional feature space. T and U are (np) matrixes of the p extracted score vectors (components, latent vectors).

Biologically Inspired Features [36] are selected to study the performance of the linear and nonlinear PLS models due to their advanced ability to express features for facial aging pattern representation. That ability is also exploited and improved in [38] by implementing a transformation to a lower dimensionality space adopting CCA based methods. Mean Absolute Error (MAE) on age estimation and accuracy (%) on gender estimation and ethnicity classification for MORPH-II Dataset can be evaluated at Table 2. KCCA [87] is the method with the best performance, and rCCA [98] (regularized version of CCA) is the solution proposed by the authors because of its good ratio between efficacy and efficiency. Other works, like [37] or [39] leaded by Guo, study the relations among the feature spaces defined by the main three humans properties extracted from their faces. The first one demonstrates that age estimation of subjects from different ethnicities can be solved with an appropriate learning-based method. The second one claims that ethnicity classification can have high accuracy under variations of age and gender, but an interesting phenomenon is observed in which the ethnic classification accuracies could be reduced by 6–8% in average in the cross-gender case of female used for training and male for testing.

3.2. Shop surveillance

The general method to investigate how shoppers behave inside a store includes following steps: modeling of environments, motion detection, human identification, tracking and behavior understanding.

Table 3
Res. of tracking algorithms from [96].

Method	CAV1	CAV2	CAV3	MFPS
STR	++	+		8
TLD	++	–	0	18
IVT	++	–	0	18
MILBOOST	++	0		12
FragTrack	0	+		8

3.2.1. Human tracking

Visual tracking algorithms are still nowadays an active area of research on the computer vision community like Smeulders et al. have exposed at their reach survey [97]. The main difficulties of the tested algorithms are illumination changes, occlusion, clutter, camera motion, low contrast, specularities and another parameters posted on [27]. The main conclusion of the survey is that the performance of 19 tested trackers on Amsterdam Library of Ordinary Videos for tracking (ALOV++[70]) is highly dependent on the scenario and the target moving object or person. Overall, Structured output Tracking with kernels (STR) [41] performs the best. Its solid performance is attributed to the use of the structured SVM (S-SVM) that accepts form of appearance training data of the detection window, although its performance is lower on scale changes of the target. Tracking, Learning and Detection (TLD) [43] is the second remarked algorithm, because its number of outstanding performances in occlusion and camera motion is much larger than any other tracker. Similar conclusions are obtained by Salti et al. [96] where different trackers that perform appearance model adaptation (like IVT [95], MILBOOST [79] or FragTrack [23]), and an evaluation methodology that enables simultaneous analysis of them is introduced. Results can be observed at Table 3. The symbols from to ++ indicate progressively better performance. 0 indicates that in general performance is not satisfactory although the algorithm is able to track the target if low overlaps or high variations on the results are acceptable. Legend: CAV1: OneStopMoveNoEnter1Cor from [71]; CAV2: OneStopEnter2Front from [71]; CAV3: ThreePastShop2Cor from [71]; MFPS: Median Frames Per Second.

In these days human tracking for indoor environments can be performed by fusion of signal processing from different kind of sensors. This topic was introduced in [88] where these techniques are classified in: early fusion (at the feature level), intermediate fusion and late fusion (at the high semantic level). An interesting combination for the feature level is proposed in [100] where electronic and visual signals are combined based on localization descriptors. Matching problem is formulated in the following manner:

$$\underset{\pi_i}{argmin} \sum_{i=1}^{n} \|x_i - y_{\pi_i}\| \tag{9}$$

The problem can be understood as they try to find the permutation of y_i to match x_i first. Where x_i is the electronic location descriptor, y_i is the visual location descriptor and π_i is the rearrangement of the series $(1, 2, \ldots, n)$. This work has been improved by Zhai et al. [67] where a multi-camera system is proposed to handle occlusions, and real-time performance is achieved with an approach that matches visual and motion signals based on location proximity, which highly improves efficiency of the system, without any loss of accuracy. Another remarkable contribution of this work is an appearance-free visual tracking, only based on human shape properties. In this case fusion is implemented with a threshold learnt offline for the following fused probability:

$$p_{i,j} = q_{i,j}^{\frac{\sigma_v^2}{\sigma_v^2+\sigma_m^2}} \cdot r_{i,j}^{\frac{\sigma_m^2}{\sigma_v^2+\sigma_m^2}} \tag{10}$$

Table 4
Evaluation of machine learning algorithms from [19].

Method	Initiator	Influencer	User	Decider	Purchaser	Passive Infl.
NB	0.679	0.735	0.606	0.614	0.614	0.882
kNN	0.659	0.740	0.630	0.603	0.651	0.865
SVM	0.692	0.748	0.728	0.599	0.724	0.910
RF	0.651	0.724	0.611	0.635	0.653	0.861

Where $q_{i,j}$ and $r_{i,j}$ are matching probabilities of detected people based on visual and motion estimations respectively. σ_v and σ_m are the standard deviations for visual and motion estimations respectively.

In [35] intermediate fusion is achieved through RFID and a computer vision tracking system based on Mean-Shift Tracker [28] to mitigate the limited tag localization capabilities of current RFID deployments. An assignment matrix is built in order to combine both sources of information. Based on the defined state-space, a Hidden Markov Model (HMM) [94] is employed to implement a probabilistic tag location. The last remarkable work in this field [25] describes a high level fusion in which a wireless sensor network is governated through co-operative Extended Kalman Filters with indirect mapping [49]. Data fusion is executed by using neural activation techniques, providing High-level behavioral computer visual monitoring and analysis, and captured detailed Wireless sensors of moving object.

3.2.2. Behavior and trajectory analysis

Once the location of the shoppers is estimated inferring its temporal correlation, a higher level method should be applied to analyze its behavior. This survey [81] enumerates the next subareas as the most relevant for this area of research: human detection, already explored in the previous subsection, activity analysis, that will be analyzed in this subsection, and interaction analysis, that will be discussed in the next subsection. The first remarkable work using an emergent technology is related to RGB-D devices [65], exposing an ability to better predict future locations of shoppers related to changes in the distribution of the products of the shops. Makela et al. [16] propose a customer behavior tracking solution based on 3D data. Experiments on retail establishments are described, classifying shopping behavior into three classes (passers-by, decisive customers and exploratory customers) with 80% accuracy. The following aspects are extracted from behavior of groups: audience structure, flow pattern and scene character. In [19] machine learning methods are tested on real-world digital signage data to predict consumer behavior. Best performance is achieved by SVM, and solution application is oriented towards the purchase decision process. Most relevant results obtained are shown in Table 4, where six roles of shoppers in a group of people are estimated. NB is the Naive Bayes Classifier [12] and kNN k-Nearest Neighbor Classifier [89]. In the same way, RFID tags are used in [47] located at the shopping cart. Tracking is performed by the information provided by the tags, and the frequent path patterns will be used to build an optimization model. Movements of customers at POS have been studied in works like explained by authors in [46]. It proposes the calculation of the following values for every module of the store: speed, stops, duration of stay and number of visits. Simple Markov [17] chains are desired to model the transitions among those modules. A new index is introduced: see-buy rate, which refers to an approximate probability to purchase a specified commodity.

3.2.3. Human-product interaction

Relations between shoppers and the products offered at a retail environment are very valuable information in order to extract the attraction of the consumers by the shop distribution and the visuality of the products. Liciotti et al. [14] propose a multicamera architecture with RGB-D sensors. When customers reach a monitored zone it is considered as an interaction that can be classified as the following states: positive(the product is picked up from the shelf), negative (the product is taken and then repositioned on the shelf) and neutral (the hand exceeds the threshold without taking anything). A complex infrastructure of wireless embedded sensors is deployed in a store to acquire human-product interaction in the work implemented by Pierdicca et al. [56]. The proposed system architecture consists of an active sensor network which can be arranged inside the store, and a series of smaller beacons which can be attached to the objects of the store. The main outputs of the system are: total number of people, average visiting time, frequency of visiting for each area of the store, number of people passing by, average group number and number of interactions per person. An approach based on RFID sensors is proposed in [52] whereas real-time human-object interaction detection is performed. In the proposed scenario, for each inventory round r, n antennas are sequentially activated t seconds, eventually returning samples S with timestamp p from the tagged population. Three events are used as feature vectors in solutions analyzed: x_1 (Received Signal Strength Indication, RSSI), x_4 ($\|\Delta mboxRFP\|$, Radio Frequency Phase difference) and x_5(Number of antennas), to build a Bayesian network whose results are compared in Table 5. An extensive experiment was performed in [57] to evaluate the feasibility of stable landmarks in a retail environment, and to demonstrate its utility in the development of next generation apps. A clustering algorithm is implemented to perform non-intuitive feature combination of sensors like Accelerometer, Gyroscope, Magnetometer, Light, Sound, Wi-Fi, GSM signal strength etc. Heterogeneity on different parameters like people (walking style), time of the day and different devices has been tested, concluding that order of influence is the same as it has been exposed in a decreasing sense.

Table 5
Evaluation of interaction algorithms from [52].

Method	Precision	Recall	F-Score	Accuracy
x_1	0.740	0.901	0.813	0.792
x_4	0.837	0.242	0.375	0.597
x_5	0.851	0.565	0.683	0.737
$x_{1,4}$	0.738	0.925	0.821	0.798
$x_{4,5}$	0.857	0.650	0.739	0.771
$x_{1,5}$	0.723	0.972	0.829	0.799
$x_{1,4,5}$	0.742	0.972	0.841	0.817

4. Conclusions

This work proposes a workflow to automatically enhance experiences of the consumers at retail establishments. Main conclusions are:

1. Face detection, normalization and expression inference: BNs is the machine learning technique offering the best results, and ASM are the most reliable to fit the shape of human face. CNN (or similar) could be a good statistical approach in order to outperform methods in the literature.
2. Head pose estimation and gaze analysis: Both fields should be implemented together in order to estimate properly attention of the subject, and feed the results of each other. Usage of RGB-

D sensors is recommended by the literature explored in this field, and CNN offers favorable results.

3. Demography: Human properties such as age, gender or ethnics should be integrated in one framework. BIF is the feature extraction technique that offers the best performance. Regarding statistical approaches the ones based on CCA, and the ones based on MMN present the most promising results.
4. Tracking: If the architecture proposed is only based on RGB sensors, STR is the most robust solution nowadays, and TLD could offer a better performance in some specific scenarios since it is more robust to scale variance. Hybrid solutions including sensors of different natures have not been deeply explored yet, but could improve the accuracy of previously mentioned algorithms.
5. Behavior analysis: Results obtained are highly dependent on the architecture proposed, algorithms only based on visual features seem to be already outperformed by those including new generation sensors. Purchase decision is the most challenging target, being SVM the best statistical solution, although this should be better verified.
6. Human-product interaction: Technologies based on the signal of smart mobile devices are required, and also a complex wireless architecture to make an analysis of a large and complex area. Mixture of high and low level features from mobile devices signal acquisition, like RFID, have been tested with satisfactory results. RGB-D sensors could cover a smaller area with comparable results. Both of them should rely on a statistical approach, like BNs or CNN, with an adapted method to reduce the dimensionality of the feature space.

We would like to acknowledge Fractalia for its support in the study developed.

References

[1] M. De Berg, et al., Delaunay triangulation, in: Computational Geometry: Algorithms and Applications, Springer-Verlag, 2008, pp. 191–218. ch. 9

[2] L. Breiman, Random forests, in: Machine Learning, vol. 45, Springer, 2001, pp. 5–32.

[3] Carcagnì, et al., Features descriptors for demographic estimation: a comparative study, in: Lecture Notes in Computer Science, Video Analytics for Audience Measurement, vol. 8811, Springer, 2014, pp. 66–85.

[4] G.M. Farinella, et al., Face re-identification for digital signage applications, in: Lecture Notes in Computer Science, Video Analytics for Audience Measurement, vol. 8811, Springer, 2014, pp. 86–96.

[5] C. Fernndez, I. Huerta, Prati, A comparative evaluation of regression learning algorithms for facial age estimation, in: Lecture Notes in Computer Science, Face and Facial Expression Recognition from Real World Videos, vol. 8912, Springer, 2014, pp. 133–144.

[6] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, in: Machine Learning, vol. 29, Springer, 1997, pp. 131–163.

[7] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, in: Analytical Chimica Acta, vol. 185, Springer, 1986, pp. 1–17.

[8] T.G. Kaya, E. Firat, Comparison of facial alignment techniques: With test results on gender classification task, in: Lecture Notes in Computer Science, Video Analytics for Audience Measurement, vol. 8811, Springer, 2014, pp. 86–96.

[9] T.G. Kaya, E. Firat, Multi-view face detection with one classifier for video analytics systems, in: Lecture Notes in Computer Science, Video Analytics for Audience Measurement, vol. 8811, Springer, 2014, pp. 97–108.

[10] V. Khryashchev, et al., Online audience measurement system based on machine learning techniques, in: Lecture Notes in Computer Science, Video Analytics for Audience Measurement., vol. 8811, Springer, 2014, pp. 111–122.

[11] A. Krizhevsky, et al., Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, vol. 25, Curran Associates, Inc., 2012. 1907–1105

[12] D.D. Lewis, Naive (bayes) at forty: The independence assumption in information retrieval, in: Lecture Notes in Computer Science, vol. 1398, Springer, 1998, pp. 4–15.

[13] S. Liao, et al., Learning multi-scale block local binary patterns for face recognition, in: Lecture Notes in Computer Science, Advances in Biometrics, 4642, Springer, 2007, pp. 828–837.

[14] D. Liciotti, et al., Shopper analytics: A customer activity recognition system using a distributed RGB-d camera network, in: Lecture Notes in Computer Science, Video Analytics for Audience Measurement., vol. 8811, Springer, 2014, pp. 146–157.

[15] H. Lodhi, Y. Yamanishi, Nonlinear partial least squares: An overview, in: Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques, ACCM, IGI Global, 2011, pp. 169–189.

[16] S. Makela, et al., Shopper behaviour analysis based on 3d situation awareness information, in: Lecture Notes in Computer Science, Video Analytics for Audience Measurement, vol. 8811, Springer, 2014, pp. 134–145.

[17] A.A. Markov, Extension of the limit theorems of probability theory to a sum of variables connected in a chain, in: Dynamic Probabilistic System, vol. 1, John Wiley and Sons, 1972.

[18] G. Rtsch, T. Onoda, K.R. Mller, Soft margins for adaboost, in: Machine Learning, vol. 2, Springer, 2001, pp. 287–320.

[19] R. Ravnik, et al., Modelling in-store consumer behaviour using machine learning and digital signage audience measurement data, in: Lecture Notes in Computer Science, Video Analytics for Audience Measurement., vol. 8811, Springer, 2014, pp. 123–133.

[20] H. Rue, L. Held, Gaussian markov random fields: Theory and applications, in: Chapman and Hall, vol. 104, London, U.K., 2005.

[21] A.J. Smola, B. Schlkopf, A tutorial on support vector regression, in: Statistics and Computing, vol. 14, Springer, 2004, pp. 199–222.

[22] H. William, et al., Support vector machines, in: Numerical Recipes: The Art of Scientific Computing, 3rd ed., New York: Cambridge University Press, 1995, pp. 883–889. ch. 16, sec. 5

[23] A. Adam, E. Rivlin, I. Shimshoni, Robust fragments-based tracking using the integral histogram, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, New York (USA), CVPR, 1, 2006, pp. 798–805.

[24] L. Brown, 3d head tracking using motion adaptive texture-mapping, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Chicago (USA), CVPR, 2001, pp. 4571–4576.

[25] Z. Chaezko, A. Kale, C. Chiu, Intelligent health care - a motion analysis system for health partitioners, in: Intelligent Sensors, Sensor Networks and Information Processing, Sixth International Conference on, Brisbane (Australia), ISSNIP, 2010, pp. 303–308.

[26] J. Chen, et al., WLD: A robust local image descriptor, in: Multimedia & IEEE Transaction Pattern Analysis on Machine Intelligent, Viena (Austria), IWSSIP, 2012, pp. 417–420.

[27] D.M. Chu, A.W.M. Smeulders, Thirteen hard cases in visual tracking, in: Advanced Video and Signal Based Surveillance, 2010 Seventh IEEE International Conference on, Boston (USA), AVSS, 2010, pp. 103–110.

[28] D. Comaniciu, V. Ramesh, P. Meer, Real-time tracking of non- rigid objects using mean shift, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Hilton Head (USA), CVPR, vol. 2, 2000, pp. 142–149.

[29] J. Connell, et al., Retail video analytics: An overview and survey, in: Proceedings of SPIE Video Surveillance and Transportation Imaging Applications, Burlingame (USA), Invited paper, 2013.

[30] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Computer Vision Pattern Recognition, IEEE Computer Society Conference on, S. Diego (USA), CVPR, 2005, pp. 886–893.

[31] A. Dhall, et al., Static facial expressions in tough conditions: Data, evaluation protocol and benchmark, in: First IEEE International Workshop on Benchmarking Facial Image Analysis Technologies BeFIT, IEEE International Conference on Computer Vision, Barcelona (Spain), ICCV, 2011.

[32] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, Palo Alto (USA), 1978.

[33] A. Ess, et al., A mobile vision system for robust multi-person tracking, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Providence (USA), CVPR, 2008, pp. 1–8.

[34] G. Fanelli, J. Gall, L.V. Gool, Real time head pose estimation with random regression forests, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Providence (USA), CVPR, 2011, pp. 617–624.

[35] M. Goller, C. Feichtenhofer, A. Pinz, Fusing RFID and computer vision for probabilistic tag localization, in: RFID (IEEE RFID), 2014 IEEE International Conference on, Orlando (USA), IEEE RFID, 2014, pp. 89–96.

[36] G. Guo, et al., Human age estimation using bioinspired features, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Miami (USA), CVPR, 2009, pp. 112–119.

[37] G. Guo, et al., A study of large-scale ethnicity estimation with gender and age variations, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, San Francisco (USA), CVPR, 2010, pp. 79–86.

[38] G. Guo, G. Mu, Joint estimation of age, gender and ethnicity: CCA vs. PLS, in: Automatic Face and Gesture Recognition 10th IEEE International Conference and Workshops on, Shangai (China), IEEE FG, 2013, pp. 1–6.

[39] G. Guo, C. Zhang, A study on cross-population age estimation, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Columbus (USA), CVPR, 2014, pp. 4257–4263.

[40] G. Guo, G. Mu, Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Providence (USA), CVPR, 2011, pp. 617–624.

[41] S. Hare, A. Saffari, P.H.S. Torr, Struck: Structured output tracking with kernels, in: Computer Vision, IEEE International Conference on, Barcelona (Spain), ICCV, 2014, pp. 263–270.

[42] L. Jia, L. Tong, Day ahead dynamic pricing for demand response in dynamic environments, in: IEEE Conference on Decision and Control, Florence (Italy), CDC, 2013, pp. 318–324.

[43] Z. Kalal, J. Matas, K. Mikolajczyk, P-n learning: Bootstrapping binary classifiers by structural constraints, in: Computer Vision and Pattern Recognition, IEEE Conference on, San Francisco (USA), CVPR, 2010, pp. 49–56.

[44] V. Khryashchev, et al., Age estimation from face images: challenging problem for audience measurement systems, in: Open Innovations Association, 2014 16th Conference of, Oulu (Finland), FRUCT, 2014, pp. 31–37.

[45] E. Kodra, et al., From dials to facial coding: Automated detection of spontaneous facial expressions for media research, in: Automatic Face and Gesture Recognition, 10th IEEE International Conference and Workshops on, Shangai (China), IEEE FG, 2013, pp. 1–6.

[46] J. Krockel, F. Bodendorf, Customer tracking and tracing data as a basis for service innovations at the point of sale, in: SRII Global Conference, San Jose (USA), SRII, 2012, pp. 691–696.

[47] H. Li, et al., Mining paths and transactions data to improve allocating commodity shelves in supermarlket, in: Service Operations and Logistics, and Informatics, IEEE International Conference on, Suzhou (China), SOLI, 2012, pp. 102–106.

[48] Z. Liu, et al., Appearance-based gaze estimation in the wild, in: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, Shangai (China), IEEE FG, 2013, pp. 1–6.

[49] K. Low, et al., Task allocation via self-organizing swarm coalitions in distributed mobile sensor network, in: National Conference on AI, San Jose (USA), AAAI, 2004, pp. 28–33.

[50] D.G. Lowe, Object recognition from local scale-invariant features, in: The Proceeding of the Seventh IEEE International Conference on Computer Vision, Shangai (China), ICCV, vol. 2, 1999, pp. 1150–1157.

[51] P. Lucey, et al., The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, San Francisco (USA), CVPR, 2010.

[52] J. Melia-Segui, R. Pous, Human-object interaction reasoning using RFID-enabled smart shelf, in: Internet of Things, International Conference on the, Cambridge (UK), IOT, 2014, pp. 37–42.

[53] S. Milborrow, F. Nicolls, Active shape models with SIFT descriptors and MARS, in: International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Lisbon (Portugal), VISAPP, 1 (2), 2014.

[54] F. Parke, Computer generated animation of faces, in: Proceedings of ACM Annual Conference, New York (USA), ACM, 1972, pp. 451–457.

[55] F. Paysan, et al., A 3d face model for pose and illumination invariant face recognition, in: Advanced Video and Signal based Surveillance, Genova (Italy), AVSS, 2009, pp. 296–301.

[56] R. Pierdicca, et al., Low cost embedded system for increasing retail environment intelligence, in: Multimedia & Expo Workshops, IEEE International Conference on, Turin (Italy), ICMEW, 2015, pp. 1–6.

[57] S. Pradhan, et al., (stable) virtual landmarks: Spatial dropbox to enhance retail experience, in: Communication Systems and Networks, Sixth International Conference on, Bangalore (India), COMSNETS, 2014, pp. 1–8.

[58] H. Ren, Z. Li, Gender recognition using complexity-aware local features, in: Pattern Recognition, 2014 22nd International Conference on, Stockholm (Sweden), ICPR, 2014, pp. 2389–2394.

[59] K. Ricanek Jr, T. Tesafaye, MORPH: A longitudinal image database of normal adult age-progression, in: IEEE 7th International Conference on Automatic Face and Gesture Recognition, Southampton (UK), IEEE FG, 2014, pp. 341–345.

[60] V. Santarcangelo, G.M. Farinella, S. Battiato, Gender recognition: Methods, datasets and results, in: Multimedia & Expo Workshop, IEEE International Conference on, Turin (Italy), ICMEW, 2015, pp. 1–6.

[61] A. Sharifara, M.S.M. Rahim, Y. Anisi, A general review of human face detection including a study of neural networks and haar feature-based cascade classifier in face detection, in: Biometrics and Security Technologies, International Symposium on, Kuala Lumpur (Malaysia), ISBAST, 2014, pp. 73–78.

[62] Y. Sugano, Y. Matsushita, Y. Sato, Learning-by-synthesis for appearance-based 3d gaze estimation, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Columbus (USA), CVPR, 2014, pp. 1821–1828.

[63] A. Sylvain, et al., Purchase intention based model for a behavioural simulation of sale space, in: International Joint Conference on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw (Poland), IEEE/WIC/ACM, 2014, pp. 318–324.

[64] A. Torrisi, et al., Selecting discriminative CLBP patterns for age estimation, in: Multimedia & Expo Workshops, IEEE International Conference on, Turin (Italy), ICMEW, 2015, pp. 1–6.

[65] E. Vildjiounaite, et al., Next generation mobile apps, services and technologies, eighth international conference on, in: Multimedia & Expo Workshops, IEEE International Conference on, Oxford (UK), NGMAST, 2014, pp. 100–105.

[66] N.B. Zahir, R. Samad, M. Mustafa, Initial experimental results of real-time variant pose face detection and tracking system, in: Signal and Image Processing Applications, IEEE International Conference on, Malaysia, ICSIPA, 2013, pp. 264–268.

[67] Q. Zhai, et al., VM-tracking: Visual-motion sensing integration for real-time human tracking, in: IEEE International Conference on Computer Communications, Kowloon (Hong Kong), INFOCOM, 2015, pp. 711–719.

[68] X. Zhang, et al., Appearance-based gaze estimation in the wild, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Boston (USA), CVPR, 2015.

[69] X. Zhu, D. Ramanan, Face detection, pose estimation, and landmark localization in the wild, in: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Providence (USA), CVPR, 2012, pp. 2879–2886.

[70] Amsterdam Library of Ordinary Videos for tracking (ALOV++) dataset avail. from: http://crcv.ucf.edu/data/ALOV++/.

[71] Context Aware Vision using Image bas. Active Recogn. dataset avail. from: http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/.

[72] Face Recognition Grand Challenge dataset available from: http://www.nist.gov/itl/iad/ig/frgc.cfm.

[73] KinFace dataset ® specifications available from: http://www1.ece.neu.edu/~yunfu/research/Kinface/Kinface.htm.

[74] MMI Initiative facial expression database ® specifications available from: http://mmifacedb.eu/.

[75] Multipie dataset ® specificationsavailable from: http://www.multipie.org/.

[76] G.B. Huang, et al., Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, University of Massachusetts, Amherst, 2007 Technical report. 07–49.

[77] J. Ahlberg, in: An updated parameterized face, Department of Electrical Engineering, Linkping University (Sweden), 2001. Report No. LiTH-ISY-R-2326

[78] P. Arbelaez, et al., Contour detection and hierarchical image segmentation, Pattern Anal. Mach. Intell. IEEE Trans. 33 (5) (2011) 898–916.

[79] B. Babenko, M.-H. Yang, S. Belongie, Robust object tracking with online multiple instance learning, Pattern Anal. Mach. Intell. IEEE Trans. 33 (8) (2011) 1619–1632.

[80] V. Badrinarayana, E.P. Becerra, S. Madhavaram, Influence of congruity in store-attribute dimensions and self-image on purchase intentions in online stores of multichannel retailers, J. Retail. Consum. Serv. 21 (2014) 1013–1020.

[81] P.V.K. Borges, N. Conci, A. Cavallaro, Video-based human behavior understanding: a survey, Circuits Syst. Video Technol. IEEE Trans. 23 (11) (2013) 1993–2008.

[82] D. Cazzato, M. Leo, C. Distante, An investigation on the feasibility of uncalibrated and unconstrained gaze tracking for human assistive applications by using head pose estimation, Sensors 14 (5) (2014) 8363–8379.

[83] D. Cazzato, et al., A low-cost and calibration-free gaze estimator for soft biometrics: an explorative study, Pattern Recognit. Lett. 14 (5) (2015) 8363–8379.

[84] T. Cootes, et al., Active shape models, training and application, Comput. Vis. Image Underst. 61 (2001) 38–59.

[85] R. Cruz-Cano, Fast regularized canonical correlation analysis, Comput. Stat. Data Anal. 70 (2014) 88–100.

[86] S. Eleftheriadis, O. Rudovic, M. Pantic, Discriminative shared gaussian processes for multiview and view-invariant facial expression recognition, Image Process. IEEE Trans. 24 (1) (2014) 189–204.

[87] D. Hardoon, S. Szedmak, J. Shawe-Taylor, Canonical correlation analysis: an overview with application to learning methods, Neural Comput. 16 (2004) 2639–2664.

[88] A. Jaimes, N. Sebe, Multimodal human-computer interaction: a survey, Spec. Issue Vis. Hum. Comput. Interact. 108 (1–2) (2007) 116–134.

[89] J.M. Keller, M.R. Gray, J.A. Givens, A fuzzy k-nearest neighbor algorithm, Syst. Man Cybern. IEEE Trans. SMC-15 (4) (1985).

[90] K. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural Netw. 2 (3) (1989) 183–192.

[91] R. Lienhart, J. Maydt, An extended set of haar-like features for rapid object detection, Image Process. IEEE Trans. 1 (1) (2002) 900–903.

[92] B.S. Manjunath, W.Y. Ma, Texture features for browsing and retrieval of image data, Pattern Anal. Mach. Intell. 18 (8) (1996) 837–842.

[93] S. Papagiannidis, E. See-To, M. Bourlakis, Virtual test-driving: the impact of simulated products on purchase intention, J. Retail. Consum. Serv. 21 (2014) 1013–1020.

[94] L. Rabiner, B.H. Juang, An introduction to hidden markov models, ASSP Mag. IEEE 3 (1) (1986) 4–16.

[95] D.A. Ross, et al., Incremental learning for robust visual tracking, Int. J. Comput. Vis. 77 (1–3) (2008) 125–141.

[96] S. Salti, A. Cavallaro, L.D. Stefano, Adaptive appearance modeling for video tracking: survey and evaluation, Image Process. IEEE Trans. 21 (2) (2012) 4334–4348.

[97] A.W.M. Smeulders, et al., Visual tracking: an experimental survey, Pattern Anal. Mach. Intell. IEEE Trans. 36 (7) (2014) 1442–1468.

[98] L. Sun, S. Ji, J. Ye, Canonical correlation analysis for multilabel classification: a least-squares formulation, extensions, and analysis, Pattern Anal. Mach. Intell. IEEE Trans. 33 (2011) 194–200.

[99] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, Image Process. IEEE Trans. 19 (6) (2010) 1635–1650.

[100] J. Teng, et al., EV-loc: integrating ELectronic and visual signals for accurate localization, IEEE/ACM Trans. Netw. 22 (4) (2014) 1285–1296.

[101] R. Valenti, N. Sebe, T. Gevers, Combining head pose and eye location information for gaze estimation, Image Process. IEEE Trans. 21 (2) (2011) 802–815.

[102] Y. Li, et al., Simultaneous facial feature tracking and facial expression recognition, Image Process. IEEE Trans. 22 (7) (2013).

[103] H. Yu, H. Liu, Regression-based facial expression optimization, Hum. Mach. Syst. IEEE Trans. 44 (3) (2014) 386–394.

[104] H. Zhong, L. Xie, Q. Xia, Coupon incentive-based demand response: theory and case study, IEEE Trans. Power Syst. 28 (2) (2013) 1266–1276.