Multimedia Tools and Applications manuscript No.
(will be inserted by the editor)

Evaluating facial recognition services as interaction technique for recommender systems

Toon De Pessemier · Ine Coppens · Luc Martens

Received: date / Accepted: date

Abstract Recommender systems are tools and techniques to assist users in the content selection process thereby coping with the problem of information overload. For recommender systems, user authentication and feedback gathering are of crucial importance. However, the typical user authentication with username / password and feedback method with a star rating system are not user friendly and often bypassed. This article proposes an alternative method for user authentication based on facial recognition and an automatic feedback gathering method by detecting various face characteristics such as emotions. We studied the use case of video watching. Photos made with the front-facing camera of a tablet, smartphone, or smart TV are used as input of a facial recognition service. The persons in front of the screen can be identified. During video watching, implicit feedback for the video content is automatically gathered through emotion recognition, attention measurements, and behavior analysis. An evaluation with a test panel showed that the recognized emotions are correlated with the user's star ratings and that happiness can be most accurately detected. So as the main contribution, this article indicates that emotion recognition might be used as an alternative feedback mechanism for recommender systems.

Keywords Recommender system · Facial analysis · Emotion recognition · Human-computer interaction

---

T. De Pessemier - I. Coppens - L. Martens
Imec - WAVES - Ghent University
iGent - Technologiepark 126, 9052 Gent, Belgium
Tel.: +32-9-264 33 25
E-mail: toon.depessemier@ugent.be
E-mail: ine.coppens@ugent.be
E-mail: luc1.martens@ugent.be

1 Introduction

Recommendation techniques have proven their usefulness as a tool to suggest interesting content items to users in many application domains such as movies, books, and music. Recommender systems learn from the users' past content selection and their feedback for the content. Therefore, users have to be identified so that the recommendation algorithm can link the current user to his past activities and expressed preferences (through feedback for the content). This research is motivated by the fact that until today, there is no accurate and user-friendly way of identifying the users of a recommender system.

The most common implementation, a username and password for authentication, requires user-computer interaction for entering the username and password, and a cognitive effort from the user for remembering these credentials. As a result, users often decide to have their password remembered by the interface of the service or never log out, especially in the context of a video watching service. This can be a problem in case of shared devices. A tablet or a TV set, for example, is often used by multiple family members living together. As a result, the recommender system might start mixing up profiles if family members do not strictly use their own account for each watching session.

New users (guests for example), typically have no account or profile for the video service. The creation of such a new user profile requires some time and input from the user. For example, users have to specify their username, password, age, etc., while creating an account. Moreover, recommender systems have a hard job to find good recommendations for these cold-start users [33], especially if also the basic demographic data, such as age and gender, is missing. So, the first challenge of a recommender system for videos is *the identification of recurring users in front of the screen, and the profiling of new users (age/gender)* without requiring many user interactions.

An additional difficulty for recommenders is that the content is often consumed simultaneously by multiple people (e.g., a family watching together). Instead of recommendations for an individual, group recommendations should be offered in this case. These group recommendations take into account the preferences of all users in front of the screen [28].

In case of a TV in the living room, people may join and leave the watching activity while the video is playing. In the ideal case, the recommender system should keep track of who is watching at all times to create accurate user profiles. However, classic recommender systems and their interfaces are not suitable for such a dynamic situation. To capture the presence of all users, each family member should log-in when he starts watching, and log-out when he stops watching. However, manually logging-in each individual user, one by one, would be time consuming and user-unfriendly. Moreover, the recommender system might draw the wrong conclusions when somebody forgets to log-out when leaving early, while others continue to watch. So in conclusion, an extensive log-in system to capture the presence and watching behavior of each individual might be too intrusive and burdensome for the users.

Therefore in current video delivery services, recommendations are typically generated based on the profile of the individual who initiates the video session. Many services enable the creation of family profiles, but these do not take into account who is actually in front of the screen or changes in the number of spectators during the video watching. So, the second challenge of a recommender system for videos is *the generation of recommendations*, group recommendations in case multiple people are in front of the screen, and the usage of implicit feedback for these recommendations.

In many content-delivery systems, users have the opportunity to provide feedback using a star rating system (5 star and 10 star rating scales are the most common). This feedback process is typically optional and therefore often skipped by the user, especially for applications with limited human-computer interaction such as video watching in a home environment. However, these ratings are the fuel of a recommender engine, and are really needed to generate accurate recommendations.

Moreover, the feedback process suffers from the same issues as the user identification process. It is unclear who is in front of the screen. Typically, explicit feedback is not requested separately for each individual in front of the screen. Usually, one rating per device is provided and the recommender is not aware of who exactly gives this rating.

Since star rating systems are often ignored by the user, an implicit feedback system is often used as alternative. The recommender system tracks which videos are watched, stopped early, etc., and subsequently derives conclusions in terms of preferences for the videos. Again, the recommender system can only guess who is in front of the screen. So, the third challenge of a recommender system for videos is *the automatic gathering of feedback for the video content* from the users in front of the screen.

The main objective of this study is a more user-friendly and practical approach for the user-recommender interaction by using facial recognition services. More specifically, the aim is to apply facial recognition in combination with the camera of the device to identify who is in front of the screen and to derive feedback by analyzing users' emotions and behavior. In this article, we present a mobile application with video recommender that automatically logs in every viewer in front of the screen (challenge 1). The recommender system fetches their preferences to compose a dynamic group for group recommendations (challenge 2). These preferences are derived from their implicit feedback, which is gathered automatically by detecting various facial characteristics during past video watching sessions (challenge 3). A mobile Android app was developed for demonstration and evaluation on a tablet. The user authentication and implicit feedback gathering with facial recognition services were evaluated based on a dataset of photos as well as with a user experiment.

The remainder of this article is structures as follows: Section 2 provides an overview of related work. The user authentication, group recommendation process, and automatic feedback gathering are described in Section 3. An extensive evaluation of the system is provided in Section 4. We further elaborate on the results in Section 5. Finally, we draw conclusions and point to future work in Section 6.

2 Related work

Authentication is defined as the verification of claimed identification [22]. Various techniques for user authentication on smartphone have been proposed in literature. These authentication techniques can be broadly categorized into two classes: physiological biometrics based techniques and behavioral biometrics based techniques [26].

The techniques using physiological biometrics rely on static physical attributes, such as face characteristics [12], fingerprints [34], iris patterns [32], and voice characteristics [6]. Some of these techniques, such as fingerprint and voice authentication, require some direct user participation, i.e. user actions that allow to test the user's identity.

Techniques using behavioral biometrics exploit patterns in user behavioral, such as touch gestures [17], physical movements or gait [30], and GPS patterns [7]. The authentication of users is performed by identifying patterns in user behavior or invariant features of users' interactions with their smartphones. These techniques analyze sampling data from built-in sensors, such as the accelerometer, gyroscope, and magnetometer [26]. Also raw data originating from interactions with the touch screen, the main input device for smartphones, is often used to identify user specific patterns. As a drawback, these techniques can only make a decision once sufficient data are available. For applications such as video watching, user movements and interactions with the device are typically limited. Moreover, these behavioral biometrics techniques are designed for smartphones but not applicable to laptops, desktops, or smart TVs.

In this research, the focus is on facial recognition (i.e. a physiological biometrics technique) and its ability to authenticate the user but also to recognize emotions, measure the user's attention, and recognize his behavior. Face detection is defined as the technique that locates the face of a person in a photo. It is the prerequisite of all facial analysis and various approaches for the detection have been studied [8]. Facial recognition is defined as the process of matching a detected face to a person who was previously detected by the system. In the study of Yang et al., this is also called face authentication and defined as the identification of an individual in a photo [38]. The similarity between two faces of two different photos can be calculated with advanced metrics using perceptual features that can include both structure and texture [16].

In the domain of facial analysis, also a lot of work has been performed regarding face-to-sketch translation problems, i.e. the transformation from face images to corresponding sketches [5]. This work has some resemblance to face authentication, e.g., regarding feature extraction. The features are often extracted through multiple hidden layers of deep (convolutional) neural networks and contain representative information that is used to distinguish an individual [9]. Recent work in this domain proposed an adaptive curriculum learning loss (called CurricularFace) that embeds the idea of curriculum learning into the loss function to achieve a novel training strategy for deep face recognition, which mainly addresses easy samples in the early training stage and hard ones in the later stage [20].

Related to face authentication, is the analysis of faces for the purpose of automatic gender and age estimation. Automatically estimating the gender and age group of the user (child, young-adult, adult, or senior) can be useful for the initial profiling of the user or for restricting content access. In this article, various commercial services for gender and age estimation are used: Microsoft's Facial Recognition Software: Face [29], Face++ [15], and Kairos [24]. Even more recognition services exist, such as FaceReader [31], but some are rather expensive or are not available as a web service that can be queried from a mobile device. So, the first research question of this study is: *"How accurate are these commercial services for gender and age estimation in view of an initial user profile for video watching?"*

Once identified who is in front of the screen, a recommender system can offer personalized recommendations (e.g., some interesting videos). But in many cases, multimedia content is selected for consumption in group, e.g., a family that likes to watch a movie together. The recommendations should then be tailored to the entire group, to ensure maximum satisfaction of the group as a whole. The group is composed of multiple individuals, who might have conflicting preferences (e.g., parents vs. children) [11].

In the context of watching TV in group, various strategies for generating group recommendations have been analyzed and compared: a common group profile, user profile merging, and merging recommendations for individuals [39]. A common group profile can be considered as a virtual user of the system, representing all group members. Through a common group profile, users cannot evaluate content individually, since they have to give ratings or provide feedback for the group as a whole. In contrast, individual user preferences are stored and processed in the cases of user profile merging and recommendation merging. The user profile merging strategy first merges all user profiles with preferences or ratings of individuals in order to construct a common group profile with the merged preferences or ratings. Subsequently, it uses a general recommendation algorithm to generate a common recommendation list for the group according to the merged group profile. The merging recommendations strategy first generates recommendation lists for each individual user according to their respective profiles using a general recommendation algorithm. Subsequently, the recommendation lists of all group members are merged into a common recommendation list for the whole group.

User experiments showed that the user profile merging strategy was the optimal solution for the case of a TV recommender [39]. Also for the merging process, different approaches have been proposed and evaluated, but still no consensus exists about the optimal solution [28,4]. In addition, these group recommender systems typically assume that all individual users who participate in the consumption are authenticated, and each of them gives feedback on the content individually. However in most practical cases, this is not obvious. So, the second research question of this study is: *"How can we calculate (group) recommendations and how are these appreciated by the users?"*

While watching video content or using an app or service in general, attention measurements and behavior analysis can help to estimate the user's interests. For example, the head pose of a user may indicate that the user is not looking at the screen. Research in this domain has shown that head pose and facial landmarks, such as eye corners, nose tip, mouth, and chin, can accurately be detected and that this detection is important for face detection and recognizing facial expressions [37]. Facial expressions of users might reveal their feelings about the content or their experience of using the service. In the field of psychology, the relationship between distinctive patterns of the facial muscles and particular emotions have been demonstrated to be universal across different cultures [13]. The psychologists conducted experiments in which they showed still photographs of faces to people from different cultures in order to determine whether the same facial behavior would be judged as the same emotion, regardless of the observers' culture. These studies demonstrated the recognizability of different emotions (happiness, sadness, anger, fear, surprise, disgust, interest).

Based on these concepts, facial expression recognition is defined as the identification of the emotions [38] by analyzing facial characteristics such as contractions or relaxations of the muscles in the face. This recognition process is often based on deep learning approaches trained on large datasets of photos or videos [19,18]. The automatic recognition of facial expressions enables the automatic exploitation of emotions for profiling and recommendation purposes. Emotions expressed by the user during video watching are a reflection of his experience with the video. Therefore, the same three commercial services were used for facial expression recognition during video watching in this study.

Various researchers have investigated the role of emotions in recommender systems. Emotion recognition during video watching has been investigated with the goal of recovering affective tags for videos [35]. The affective state of the user can be determined with a high accuracy using electroencephalogram (EEG). But compared to taking photos during video watching as in our study, EEG is much more intrusive. A study similar to our research proved that the performance of a video recommender can be enhanced by using facial expression data such as recognized emotions [3]. They utilized facial expression data as a complementary source of information in addition to the traditional click-through data. In contrast, our study investigates if these facial expression data correlate with explicit feedback in the form of a 10 star rating mechanism. In case of a significant positive correlation, the recognized emotions can be considered as an alternative for the rating feedback. These researchers of that similar study also pointed to another interesting research track in the domain: investigate the role of complementary behavior such as the head pose [2]. This behavior is also recognized and analyzed in our study.

Emotions can be used to improve the quality of recommender systems in three different stages of the content consumption process [36]:

1. The entry stage: when a user starts to use a content delivery system with or without recommendations, he is in an affective state, the entry mood. The user's decision making process is influenced by this entry mood. A recommender system can adapt the list of recommended items to the user's entry mood by considering this as contextual information [1].
2. The consumption stage: after the user starts to consume content, he experiences affective responses that are induced by the content [36]. Moreover, by automatic emotion recognition from facial expressions, an affective profile of movie scenes can be constructed. Such an item profile structure labels changes of users' emotions through time, relative to the video timestamp [23]. We will call this item profile, an "emotion fingerprint" in this article.
3. The exit stage: after the user has finished with the content consumption, he is in the exit mood. The exit mood will influence the user's next decisions. In case that the user continues to use the content delivery system, the exit mood for the content just consumed is the entry mood for the next content to be consumed [36].

In this article the focus is on the second stage: the consumption stage. Users watch movies and their facial expressions are captured as a vector of emotions, attention measurements, and behavior values, that change over time. These values of the user's facial expressions, such as the recognized emotions, are used as an indicator of the user's satisfaction with the content. The assumption is that users appreciate a video if they sympathize with the video and express their emotions in accordance with the expected emotions.

Therefore, the third research question of this study is: *"Can facial expression recognition during video watching be used as an unobtrusive (implicit) feedback collection technique?"* According to our knowledge, this research is the first that investigates emotion recognition, attention measurements, and behavior analysis as feedback mechanism for recommender systems and the first to study how explicit ratings correlate with recognized emotions.

3 Method

This study investigates if facial recognition services can facilitate human-computer interaction for video watching services. Therefore, an Android application has been developed with the following three subsequent phases: 1) User authentication with an automated login procedure and user profiling (gender and age) based on facial recognition to identify all people who are in front of the screen. 2) Personalized recommendations (group recommendations in case multiple people are in front of the screen). 3) Automatic feedback gathering while the chosen video is playing. Using the front-facing camera of a tablet or smartphone or a camera connected to a smart TV, the app takes photos of all people in front of the screen and sends requests to different facial recognition services.

Figure 1 visualizes the data flow of the app and the used facial recognition services. In the first phase, the goal is to identify and recognize each face in the photo. For new faces, age and gender will be estimated to create an initial user profile. In the second phase, (group) recommendations will be generated for all identified users. These personalized recommendations can help users in the content selection process and demonstrate the added value of facial expression recognition. In the third phase, the photos will be used for emotion recognition, attention measurements, and behavior analysis in view of deriving automatic feedback. This feedback can be used as input for the recommender system to improve the accuracy of the (group) recommendations.

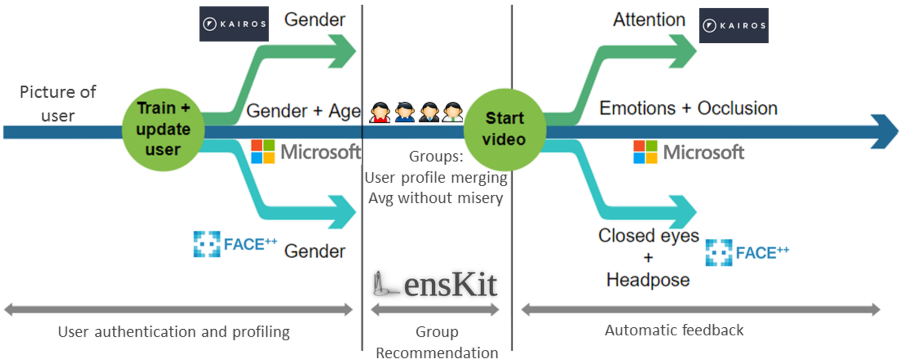

Figure 1 Data flow consisting of 3 phases: user authentication and profiling, group recommendations, automatic feedback.

3.1 Phase 1: User authentication and profiling

Facial recognition software is often used to unlock smartphones automatically using the front-facing camera. However, the applications of facial recognition in a group context, to identify multiple people simultaneously, are less common. In other words, facial recognition is used in this phase to give an answer to the question: “who is in front of the screen?”. In a real world scenario, it can be multiple people, all of whom will be individually identified.

To authenticate recurring users (who have been identified by our app in a previous session), our Android app uses Face Storage of the Microsoft Face service [29]. Identified persons are saved with their recognized face in a Person Group, which is trained based on the photos of the camera. This enables to link the user in front of the screen with one of the existing user profiles. For new users, who do not have a profile yet, the age and gender is estimated (Section 4.1). To cope with the cold-start problem, initial recommendations (Phase 2) are based on these demographics.

In practice, user authentication and profiling works as follows in our app. Users can log in automatically by ensuring that their face is visible for the front-facing camera when they push the start button. A photo is made and used as input for the facial recognition services, which will try to recognize the faces in the photo. If the recognized face is matching a recurring user, this user is logged in automatically and his existing profile (age, gender, and watching history) is retrieved for the recommender (Phase 2). In addition, the new photo is used as a new training sample for Face Storage. New users do not have a stored profile; for them a new profile is created based on their age and gender as estimated by the facial recognition services. After every login, the age estimation is adjusted based on a new estimation with a new photo. This update can correct age estimations based on previous photos, but also takes into account the aging of users when using the app for multiple years. This is especially useful for children and minors who can get access to more content as they fulfill the minimum age requirements over time. Moreover, storing a photo for every session has the advantage that changes to the user's appearance (e.g., different hairstyle) can be taken into account.

3.2 Phase 2: Group recommendations

After authentication, users receive recommendations for videos. Figure 2 shows a screenshot of the user interface with these recommendations. The group recommendations are generated by merging individual user profiles (consisting of age, gender, feedback, and watching history); one profile for every user in front of the screen. Preferences of all individual users are merged into a group profile, which is used for recommendations. From the eleven merging strategies proposed by Judith Masthoff [28], the "Average without misery" strategy was adopted in our group recommendation algorithm. This strategy takes into account the preference score of every user by calculating a group average, while avoiding misery by eliminating videos that are really hated by some group members, and therefore considered as unacceptable for the group.

The group profile is used to generate recommendations for interesting content items for the group based on the users' historical viewing activities and their preferences. The preferences are based on the users' feedback for the content. In addition, the viewing history is used to avoid recommendations for movies that users have already seen. So, if at least one of the users has already seen the video, it is considered as unsuitable for the group since this person probably does not want to see the video again. Figure 3 gives a graphical overview of the recommendation process. These recommendations based on the user profiles are indicated as "history based recommendation" in Figure 3. The recommended items are selected based on a prediction of the rating score for the items by the group. The items with the highest predicted ratings are recommended. The Lenskit [14] recommendation framework was used to calculate these rating prediction scores and convert them into a Top-N recommendation list. Lenskit is an open source toolkit that contains a set of state-of-the-art recommendation algorithms. In case only one user is in front of the screen, the process is very similar, but recommendations are generated based on this single user profile.

Besides personal preferences, other criteria, such as the age and gender of the users, are taken into account. This is indicated in Figure 3 as “demography based recommendation”. For cold-start users, i.e. new users without watching history, their influence on the recommendations is only determined by these demographics. Age is modeled as classes of age ranges, firstly to filter out inappropriate content for minors, secondly for estimating the ratings for cold-start users based on other users of the same class. We used the age ranges that are also used by IMDb: <18, 18-29, 30-44, 45+.

The age of the users is used to determine whether a video is suitable for the viewers. For every video, the advised minimum age is retrieved from the Common Sense Media website [10]. If at least one of the users is younger than this age threshold, the video is marked as unsuitable for the group according to the average without misery strategy.

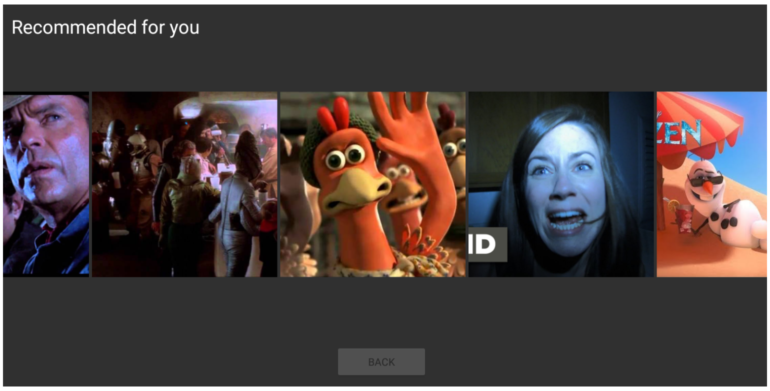

Figure 2 A screenshot of the user interface showing the recommendations.

If a new user is present in front of the screen, i.e. a cold-start user, the recommendations are only based on the user’s demographics, since his watching history is not available. An estimation of the user’s age and gender, as provided by the facial recognition services, is used to find users with similar demographics. The preferences of that demographic class (gender & age class) are used to estimate the preferences of the cold-start user. In case of an explicit rating for example, we use the mean rating of that demographic class for the movie, as mentioned by IMDb [21]. IMDb collects and aggregates ratings from their website visitors, and provides for each demographic class the results. In our recommender, the mean rating provided by the demographic class is compared with the mean rating over all users for this specific movie. This difference (demographic class mean - global mean) indicates if the movie is less or more suitable for a specific gender & age class.

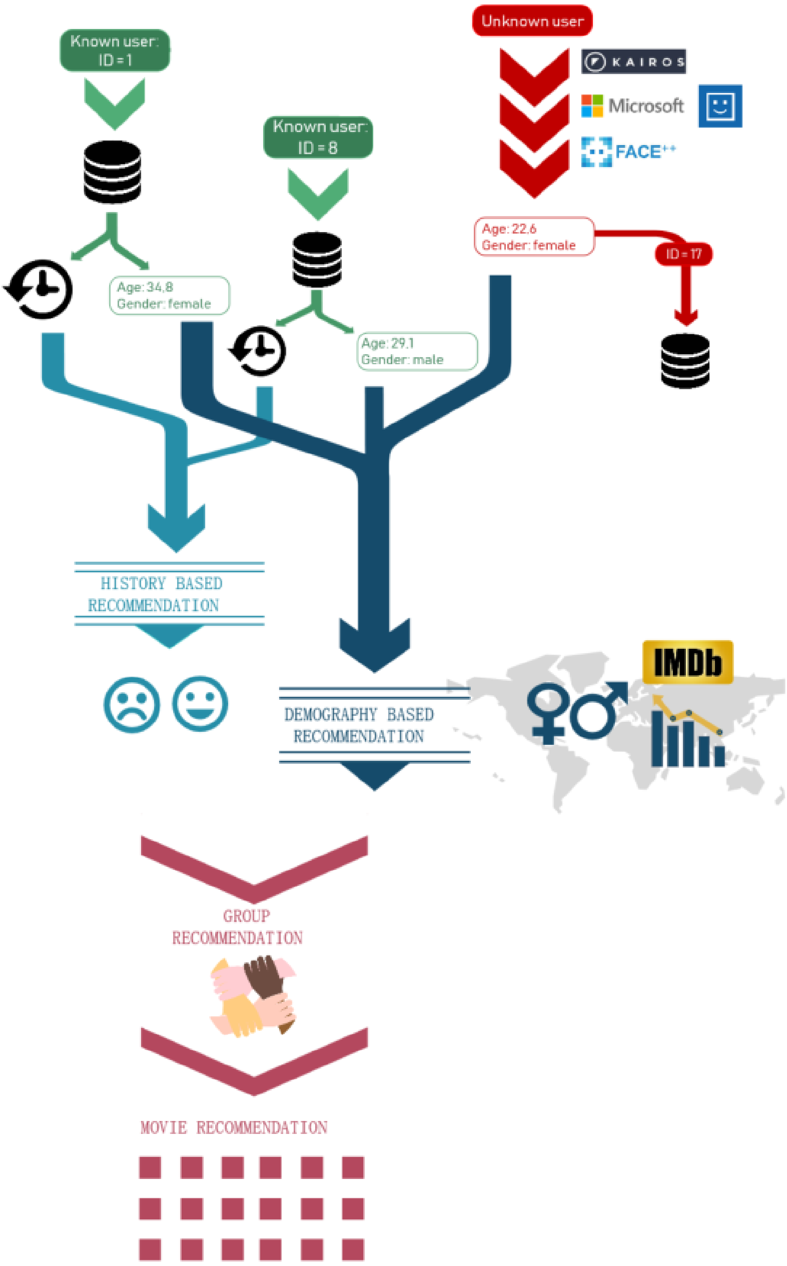

Figure 3 Graphical overview of the recommendation process: History based recommendation for users with a watching history; Demographic recommendations based on the user's demographics.

3.3 Phase 3: Automatic feedback

Commercial services that perform emotion recognition, attention measurements, and behavior analysis are often based on the analysis of photos. Therefore, our Android app continuously takes photos from the users with the front-facing camera during video watching. Every second, a photo is taken and sent to the Microsoft recognition service for face detection and authentication in order to check if all viewers are still in front of the screen (phase 1). Subsequently, for each identified face, the area of the photo containing the face is selected and the photo is cropped so that only one person's face is visible. Next, the cropped photo is sent to each of the three recognition services. Since facial recognition services are queried for every identified face with another part of the photo as input, facial expressions will be recognized for all identified individuals in front of the screen, one by one.

For recognizing emotions on the users' face, the Microsoft service was used in our app. This recognition services provides a list of values, one for each of the six possible emotions: anger, disgust, fear, happiness, sadness, and surprise. But these recognized emotions cannot be directly used as implicit feedback [3], since different videos evoke different emotions. One can assume that if users express their emotions during video watching, and these emotions are matching the expected emotions, then users might sympathize with the video and appreciate it. It is important that the expressed emotions are matching the emotions that are expected based on the content of the video. E.g., during a horror scene 'fear' can be expected, whereas during a comedy scene users may laugh ('happy' emotion). Recognized emotions that are not expected from the content, might be due to external influences (e.g., other people in the room) or reflect contempt for the video (e.g., laughing with terrifying scenes of a horror movie). Therefore, unexpected emotions are not taken into account as feedback in our recommender.

So, the similarity between the expressed emotions (=recognized emotions) and the expected emotions is calculated to determine the user's experience with the video while watching it. The expected emotions are represented by the emotion fingerprint, which is defined as a unique spectrum of expected emotions for a video scene. For every second of the video, the emotion fingerprint specifies the probability value of each of the six possible emotion dimensions: anger, disgust, fear, happiness, sadness, and surprise. These emotion dimensions have been identified in the field of psychology [13]. So, the emotion fingerprint indicates which emotions the video typically provokes among viewers at every second of the video. The emotion fingerprint is composed by aggregating emotions expressed by many users during watching this specific video. Section 4.5 explains in detail how the fingerprint of a video scene is computed based on an example. Figure 4 shows a schematic overview of the comparison between expected emotions, also called the emotion fingerprint, and the expressed emotions. These are used to derive feedback for the content automatically.

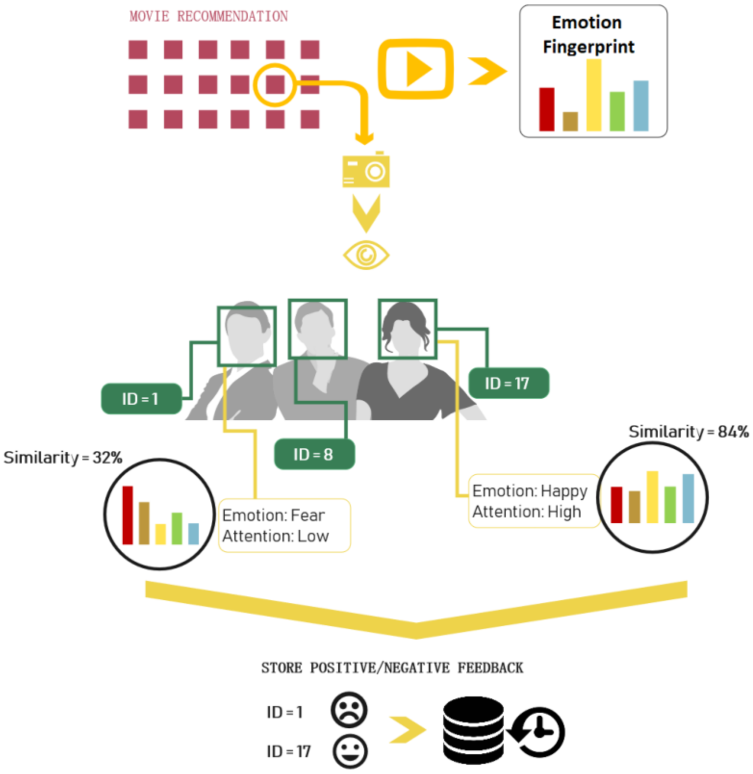

Figure 4 Graphical overview of the feedback for the recommender system: Users' expressed emotions are compared to the emotion fingerprint of the video scene.

The similarity between expressed and expected emotions is calculated based on the inverse of the emotion distance and an additional constant to avoid a division by zero.

$$emotionSimilarity = \frac{1}{emotionDistance + 1} \tag{1}$$

The emotion distance, the distance between expressed and expected emotions, is calculated based on the euclidean distance between the values of these two emotion spectra for every second i of the video and each emotion j. For the expressed emotions, the output of the Microsoft service is used in our online experiment (Section 4.5) because this service gave the best results in the offline evaluation (Section 4.4).

$$emotionDistance = \sqrt{\sum_{i=0}^{n}\sum_{j=1}^{6}(expressed_{i,j} - expected_{i,j})^2} \tag{2}$$

In addition to emotions, also the attention level and user behavior are analyzed during video watching as an additional implicit feedback mechanism.

The attention level is a value specified by the recognition service that estimates how much the user pays attention. Furthermore, three aspects of user behavior are analyzed during video watching: covering part of the face (occlusion), closing the eyes, and turning the head away from the screen. Here, the assumption is that users who are not paying attention during video playback (e.g., closed eyes, not looking at the screen), are not interested in the video.

The Kairos recognition service has a built-in feature that represents the *attention level* of the user, which is estimated based on eye tracking and head pose.

The Microsoft service has an interesting feature that recognizes occluded areas of the face, as part of the recognition service. This *occlusion* is used to recognize an intensive user experience during video watching in case users respond to the video by holding their hands in front of their mouth or eyes (typical for shocking content).

Face++ is the only one of the three recognition service that can detect *closed eyes*. Closed eyes may indicate that the user is sleeping in front of the screen, and therefore this is interpreted as a user who is not paying attention. We record this as negative feedback for the video. Also the user's *head pose* is derived from the Face++ service. Although other facial recognition services can recognize the head pose as well, the estimation of Face++ showed to be the most accurate one. In case users do not want to see a scene, they might turn their head and look away. Also this is interpreted as negative feedback.

In our Android application, these behavioral aspects are combined into the *overallAttention* level by aggregating the service results over all photos taken during video watching. The overall attention level is calculated as the percentage of photos in which the user pays attention and following conditions are met: Kairos' attention level > 0.5, no occlusion of the face, both eyes open, and head pose angles are between the margins: 20 degrees for the pitch angle and 30 degrees for the yaw angle. The assumption is that the user is not paying attention to the video if one of these conditions is not met.

$$\text{overallAttention} = \frac{\#\text{Photos}(\text{attentionLevel} > 0.5 \ \& \ \text{noOcclusion} \ \& \ \text{eyesOpen} \ \& \ \text{headPose})}{\#\text{Photos}} \tag{3}$$

An *implicitFeedbackScore* on a scale ranging from 0 to 10 is calculated by aggregating the different facial analysis features: emotions and attention. The similarity with the expected emotions has a contribution of six points out of ten points. The overall attention level counts for the remaining four points. An implicitFeedbackScore of 10 means that the user's expressed emotions are perfectly matching the expected emotions and that the user is paying full attention to the video. This implicitFeedbackScore is used as input for the recommendation algorithm. This score is assigned to the movie that the user was watching, and it specifies the user's preference for this movie, just as a traditional star rating.

$$\text{implicitFeedbackScore} = 6 \cdot \text{emotionSimilarity} + 4 \cdot \text{overallAttention} \tag{4}$$

4 Evaluation

Evaluations of commercial facial recognition services have been performed in literature, but are typically based on datasets with high-quality photos that enable an accurate recognition: high resolution, sufficiently illuminated, no shadow or reflections, and a perfect position of the face in the middle of the photo [12,3]. In contrast, for facial recognition and analysis during (mobile) video watching, the front-facing camera of the device is used without flash, which yields not always ideal photos.

Therefore, we evaluated the three facial recognition services in an offline test (based on a publicly available dataset of photos) in Section 4.4 as well as in an online setting (with real users using our app). For the evaluation of the age and gender estimation (Section 4.1), 46 users ranging from 0 to 66 years old were involved in our test. For the evaluation of the attention level (Section 4.3), we used 76 photos of our test users with different levels of attention. The evaluation of emotion recognition during video playback (Section 4.5) requires more time from the user and was therefore performed by only 20 users. Based on the evaluations of the video, these 20 users received recommendations. Through a questionnaire, users could evaluate these recommendations and the implicit feedback technique, as an alternative for ratings (Section 4.2).

The overall aim of this study is to improve the user friendliness of recommender systems for video watching in the living room. This evaluation is the first step to reach the future goal of multi-user recognition and is therefore carried out with a tablet, in a rather controlled environment, with one person at a time. During the test, photos were taken with the front-facing camera of the tablet (Samsung Galaxy tab A6). If the tablet would have captured two people in the photo, the recognition process would be performed for both recognized faces.

To have a realistic camera angle, the users were asked to hold the tablet in front of them, as they would usually do for watching a video. The users were sitting on a chair in front of a table and the room was sufficiently illuminated. However, no guidelines were provided regarding their head position, attention, or behavior; e.g., nothing was said about looking away or closing eyes. Most users were focusing on the screen during video watching. During explicit video scenes, e.g., a disgusting scene or a horror scene, some users turned their head to look away from the screen, or they closed their eyes. The photos taken with the front facing camera are used as input for the recognition services. Since all users used the app on the tablet individually, most photos clearly show the face of the user as required for the age & gender estimation and emotion recognition.

4.1 Age & gender estimation

As an important aspect of the first phase of the recommender process, the authentication was evaluated. This means recognizing the users who used the app in the past. The automatic authentication of the 46 users (login process) showed to be very accurate: 4 undetected faces with Kairos (9%), 2 with Face++ (4%), and 2 with Microsoft (4%).

Subsequently, for the recognized faces, the recognition services were used to estimate the users' age and gender based on photos of the test users taken while holding the tablet. The estimated age and gender, as provided by the recognition services, were compared to the people's real age and gender. Figure 5 visualizes the differences between estimation and real age, sorted according to the real age of the test users. The largest errors were obtained for the estimation of the age of children. Face++ and Kairos typically estimate the children to be older than they are. Table 1 reports the number of photos for which a detection was not possible, the average absolute age error, the median age error, and the percentage of photos for which the gender estimation was wrong.

The results of the three facial recognition services were combined into a hybrid solution that aggregates the results of the three. For the age estimation, the aggregation is the average of the three age estimations of the recognition services. For a decision on the gender, the gender estimations of the three services are aggregated using a voting process. For each photo, the gender with the most votes (2 or 3 out of 3) is the result of the aggregation. This voting aggregation showed to be more reliable than each individual service for estimating gender. Microsoft turned out to be the best in the age test, so we decided to use only this service to estimate the user's age. For the gender estimation, the aggregation method is used.

So as an answer to the first research question, we can state that the facial recognition services provide an accurate age and gender estimation for creating an initial profile for a cold-start user.

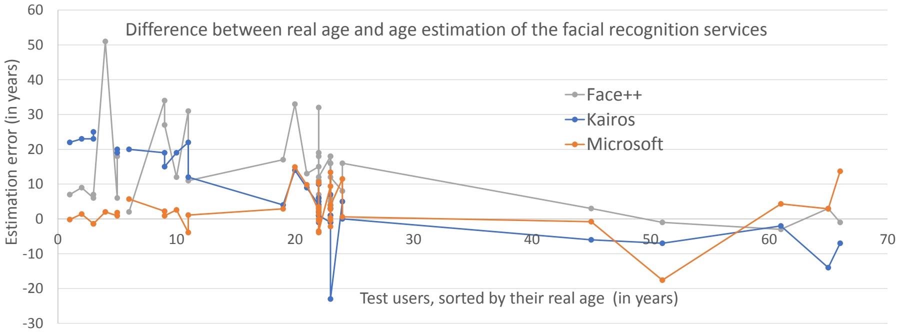

Figure 5 Evaluation of the age estimation using facial recognition services.

4.2 Recommendations

After watching the videos (of the experiment of Section 4.5), the 20 test users were asked to answer a questionnaire that assesses how good the recommend-

Table 1 Evaluation of gender & age estimation using facial recognition services.

	Face++	Kairos	Microsoft	Aggregation
Detection failed	2	4	2	2
Avg. abs. age error	13.14	8.88	4.31	7.91
Median age error	11.0	6.0	2.9	8.1
Gender error (%)	13.6	11.9	15.9	11.3

ations match their interests, and how they typically express their emotions during video playback. This questionnaire was not part of the app, but offered as a separate form to make it clear to users that the only goal of the questionnaire is the evaluation in the context of the experiment.

Table 2 lists the most important questions and their answers. Two test users did not participated in the questionnaire. One other test user did not answer the question regarding the usage of the app to give implicit feedback to video content. So, only 17 votes are collected for the last question, the other questions received 18 votes.

Question 1 shows that the majority of the users feels that the offered recommendations match their personal preferences. So, the recommender system is able to generate accurate recommendations based on the gathered implicit feedback. Questions 2 and 3 ask if users express their emotions during video watching. About half of the users indicate that they express emotions (Question 2), but Question 3 shows that not all emotions are expressed in the same way. According to the users, they express rather emotions such as happiness and disgust, than anger or fear. Happiness is often expressed through a smile for example, whereas fear is subtly expressed through the eyes.

Question 4 asks how people react to movie scenes they would rather not like to see (e.g., explicit horror scenes). Most people look away; some close their eyes; a minority covers their eyes or face. Some of the test users say that they sometimes hold their hand(s) before the mouth, in case of emotional scenes (e.g., surprising or sad movie scenes).

The answers of Question 5 show that most of the users prefer to continue to use this app with automatic implicit feedback as an alternative for explicit star ratings. The users who answered “no” to this question, typically have some concerns regarding the photos that are taken during video watching and their privacy.

So as an answer to the second research question, we conclude that for most users the recommendations based on implicit feedback match their preferences. However, not everyone expresses his emotions in the same manner, and some users say that they barely express emotions during video watching.

4.3 Attention level offline

The features that constitute the overall attention score of the user (equation 3) were evaluated with a dataset that we created with photos of the users taken

Table 2 User questionnaire about recommendations and implicit feedback.

No	Question	Yes votes	No votes
1	Are the offered recommendations suitable for you and matching your movie preferences?	15	3
2	Do you feel that you have shown many emotions when watching the movie scenes?	11	7
3	Do you think that you would express the following emotions while watching movie scenes or movies in general (even slightly)?		
	Anger	3	15
	Fear	9	9
	Happiness	15	3
	Surprise	10	8
	Sadness	11	7
	Disgust	14	4
4	Do you think that you would react to (very explicit) movie scenes in the following ways?		
	Looking away	14	4
	Closing your eyes	6	12
	Covering your eyes	3	15
	Covering your face	3	15
	Holding your hand(s) in front of the mouth	4	14
5	Would you continue to use this app with implicit feedback to get better and more personalized recommendations, so that you do not have to give explicit feedback?	11	6

during the test. In addition, we included some photos where users were explicitly asked to cover part of their face. The photos were manually annotated with the feature labels (e.g., eyes closed or not) to obtain the ground truth. The result was a dataset of 76 photos with a focus on these attention features (e.g., multiple users covering their mouth, eyes, etc.).

Table 3 lists the percentage correctly recognized photos for each attention feature. However, not all attention features are available for the three services. Features that are not available are indicated with N/A.

Face++ provides two probability values for closed eyes (for left and right eye). If both values have a probability of 40% or more, we consider this as “closed eyes”.

Kairos estimates the attention level of the user and expresses this with a value between 0 and 1. Kairos’ attention feature is based on eye tracking and head pose. To convert this to a binary value (attention or not), we used a threshold of 0.5.

Kairos and Face++ can recognize the head pose of the user. If the head position is outside the boundaries (20 degrees for the pitch angle and 30 degrees for the yaw angle), we interpret this as “head turned away and not paying attention”. The estimation of Face++ is more accurate than this of Kairos. Therefore, the head pose specified by Face++ is used in the app.

If the face is turned away too much from the camera (e.g., if a user is looking in the opposite direction) or if a large part of the face is covered, then face detection might fail. The percentage of “no detections” is also indicated in

Table 3 Evaluation of the attention level: percentage correctly recognized

	Face++	Kairos	Microsoft
Covering eyes	N/A	N/A	97.37%
Covering mouth	N/A	N/A	94.74%
Covering forehead	N/A	N/A	98.68%
Closed eyes	97.37%	N/A	N/A
Attention	N/A	82.97%	N/A
Head pose attention	72.37%	60.53%	N/A
No detection: Face turned away	2.36%	11.84%	7.89%

Table 3. Remember that this dataset was created with the focus on attention level. For many photos, users were explicitly asked to turn their head away. Therefore, the number of "no detections" is rather high.

4.4 Emotion recognition offline

The emotion recognition ability of the three facial recognition services was evaluated using the Cohn Kanade dataset [25, 27], a publicly available dataset which contains photos of people expressing different emotions ranging from a neutral face expression to a very explicit expression of the emotion. So, the emotion that the person is expressing in the photo is specified in the dataset as ground truth. Six photo sets, containing the photos with the very explicit emotion expressions (one set for each emotion), are used as input for the facial recognition services. The output of the recognition services is a vector of 6 values, one value for each emotion. For Face++ and Kairos, these output values range from 0 (meaning this emotion has not been recognized at all) to 100 (meaning this emotion has been recognized with great certainty). For the Microsoft service, the output values range from 0 to 1 (with the same interpretation). For evaluation, these outputs of the recognition service are compared to the emotion labels of the dataset (ground truth).

Figure 6 shows for each of the six photo sets how the emotions are recognized by the three services. The emotion values are shown on the Y-axis for each photo set that was used as input (photo index on the X-axis). Each recognized emotion has a different color. For a specific photo set, the ideal emotion recognition should result in the detection of only one emotion with a value of 1 for Microsoft and 100 for Face++ and Kairos, while the other emotion values are 0. For example, for the graphs that have "surprise" in the title, only photos of surprised people are used and the perfect recognition should only detect "surprise". So, a maximum value should be obtained for the yellow line (surprise). For a few photos, the person's face could not be detected. This resulted in no output of the service. Therefore, not all indices have an emotion value in the graphs of Kairos.

In general, the results clearly show that some emotions, such as surprise and happiness, are more easy to detect with a high certainty, whereas other emotions, such as fear, are more difficult to detect and can easily be confused.

Although the people of these photos are expressively showing their emotions, the automatic recognition of these emotions by professional facial recognition services is not yet perfect.

Anger is accurately recognized by Kairos and Microsoft, whereas Face++ confuses anger with disgust and sadness for some photos. Fear is the most difficult to detect: Kairos detects fear in most photos; but Face++ and Microsoft sometimes incorrectly recognize sadness and disgust. Happiness is very accurately detected by all three services. The results obtained with the Microsoft service are almost perfect: only happiness is detected and no other emotions. Also surprise is very well recognized by all three service with high emotion values. Sadness is recognized for most photos, but in comparison to happiness and surprise, the emotion values are lower. This indicates that sadness is less clearly recognizable for emotion recognition services. Disgust is sometimes confused with anger, but Face++ and Microsoft rightly assign a much lower value to anger for most photos.

In conclusion, the comparison between the recognized emotions and the true emotion labels of the photos, revealed that the Microsoft service has the most accurate emotion recognition. Therefore, the Microsoft service was chosen as solution for emotion recognition in Section 4.5. The evaluation using the Cohn Kanade dataset (Figure 6) also indicated that - even with the most explicit emotion photos - anger, disgust, and fear are always detected with a low probability value. Happiness can be detected with high probability values. So, happiness can be considered as the emotion that is rather easy to detect with a high confidence, whereas fear, anger, and disgust are much harder to detect.

4.5 Emotion recognition online

Emotion recognition as a tool for gathering automatic feedback, was evaluated with a test panel consisting of 20 users between the ages of 5 and 72. During the test, each user watched six videos with a duration of 3 to 5 minutes on a tablet. For each of the six basic emotions, one characteristic video was chosen (e.g., for happiness a comedy, for fear a scary horror movie, etc.). During video watching, the front-facing camera continuously took photos, which were analyzed and for which an emotion score (based on equation 2 and 1), overall attention score (equation 3), and implicit feedback score based on a complete facial analysis (equation 4) were calculated. After each video, the users were asked to express their preferences for the video using a 10 star rating mechanism in order to compare this with the calculated implicit feedback score.

The emotion fingerprint of the video was obtained by aggregating the emotions expressed by all the test users during the specific video. Figure 7 gives an example of this aggregation for a comedy video (a scene from the movie “Dude, Where’s My Car?”). The emotion signal of the fingerprint is the average emotion value over all users at every second of the video. So for every user, a photo

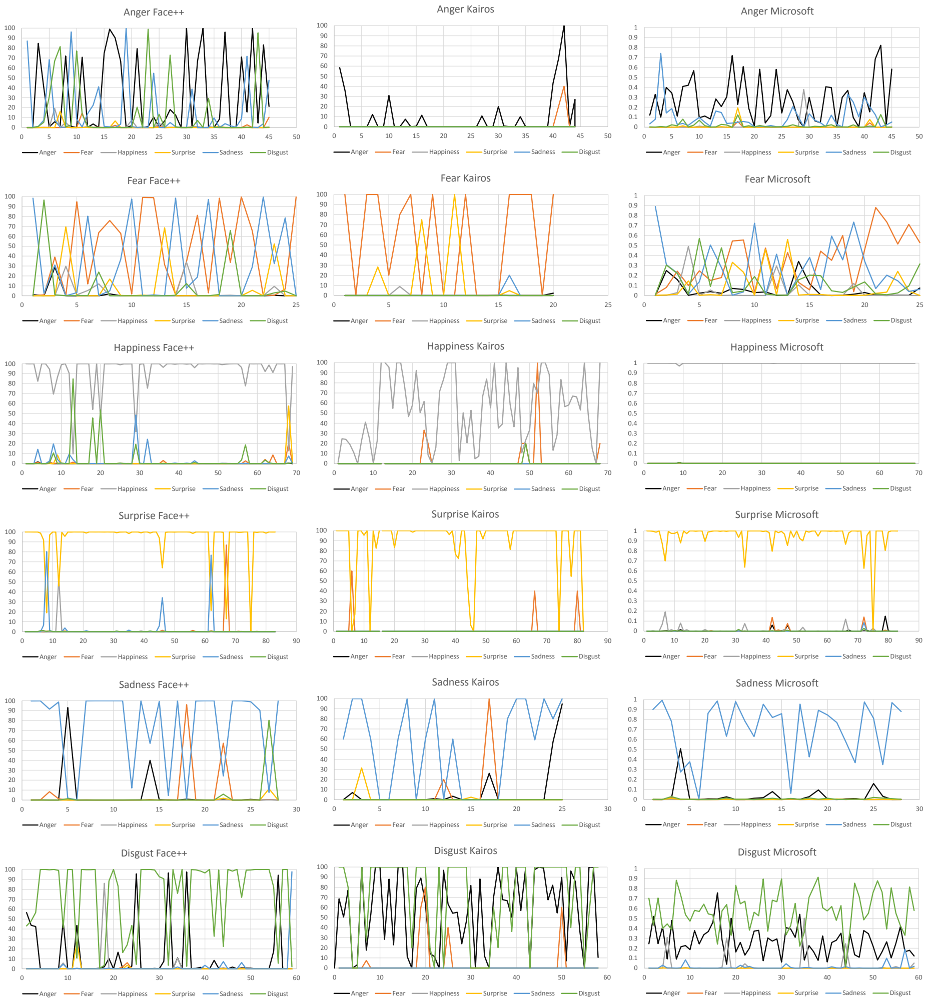

Figure 6 Output of the recognition services: recognized emotions in photos of people expressing emotions.

was taken every second while watching that video. The emotions expressed at that second were averaged over all users to obtain an emotion fingerprint that represents the emotions of all users at that second of the video. For each of the six basic emotions a probability value is obtained, which indicates the chance that that emotion will be expressed at that second of the video.

Because of this aggregation of emotion values coming from multiple test persons, the emotion fingerprint was constructed after the user test. Subsequently, irrelevant emotion values are filtered out and only the most dominant emotions are retained (e.g., happiness, surprise, and sadness in this comedy movie). Key scenes of the video that may provoke emotions are manually selec-

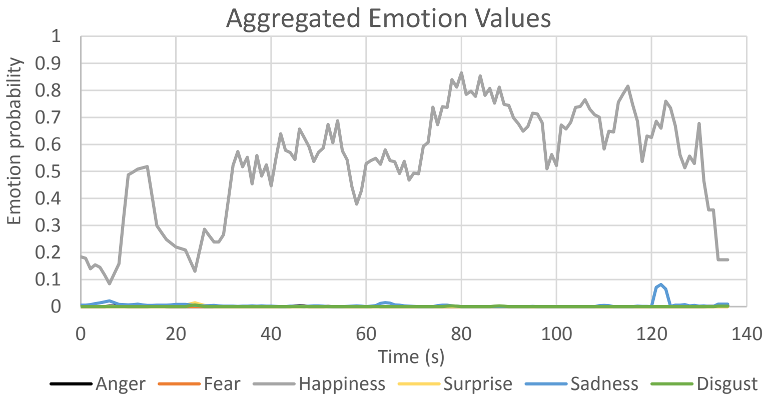

Figure 7 The emotion values aggregated over all test users.

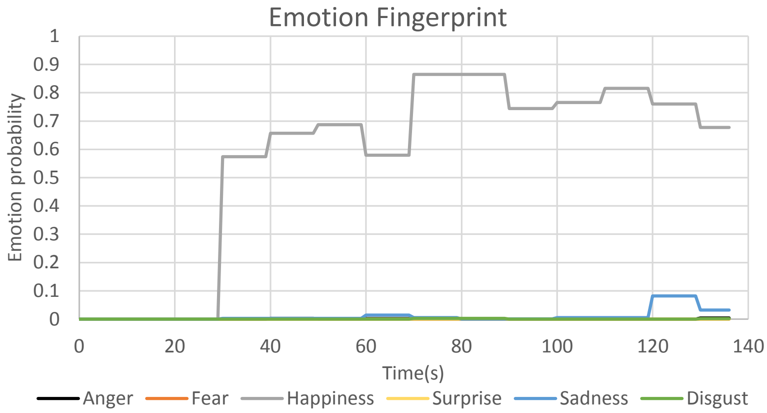

Figure 8 The emotion fingerprint based on the aggregated emotions.

ted. During periods of the video without expressive emotions, the fingerprint values are set to zero. During these periods, we assume that the emotions recognized from the users' face are due to external factors. As visible in Figure 8, the video contains no emotional scene from second 0 until 30. Next, the fluctuations of the emotion signal are reduced by using the maximum observed emotion value over a time window of 10 seconds. This time window takes into account that an expression of emotions typically takes multiple seconds. Figure 8 shows an example of a resulting emotion fingerprint for the comedy. We consider this emotion fingerprint as the expected emotion spectrum for the specific video.

To discuss the results, we elaborate on the emotion spectrum of three users of the test. Figure 9 shows the emotions expressed while watching the comedy video for three representative users with ID 3, 4, and 13. The expressed emotions of users 4 and 13 clearly show some similarities with the emotion fingerprint. Happiness is the most dominant emotion, but also some sad and

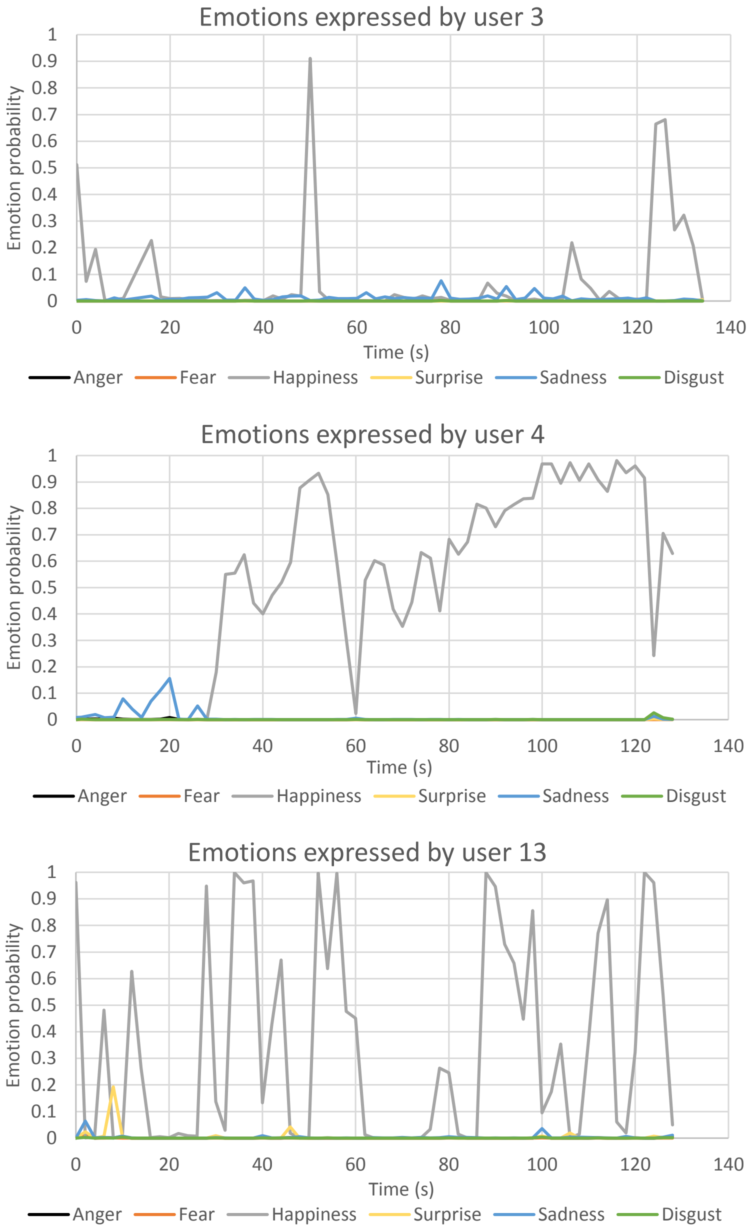

Figure 9 Emotions expressed by three users during video watching. Users 4 and 13 like the video, user 3 does not like it.

surprising aspects are in the movie. The video contains the most expressive emotions (funny scene) from second 30, which is visible in the expressed emotions of users 4 and 13. In contrast, the expressed emotions of user 3 are very different from the emotion fingerprint.

The explicit ratings of the users with ID 3, 4, and 13 were respectively: 3, 6, and 6.5 stars on a scale from 1 to 10 for this video. The low explicit rating of user 3 is reflected in the emotion values of this user (implicit feedback), which are significantly lower than with the other users. Similar results are obtained for other users. For the test with 20 users, we achieved a significant positive correlation of 0.37 between the explicit rating given by the user, and the similarity between the user's expressed emotions and the expected emotion fingerprint (equation 1).

Since the emotion recognition and rating process are characterized by a lot of noise, the correlation between both will never be perfect. However, the positive correlation indicates that expressed emotions clearly are a form of implicit feedback that can be used as input for a recommendation algorithm. Moreover, we expect that the correlation might improve if users watch full movies or tv shows instead of short movie scenes, as in our user test. Therefore, we can consider the recognized emotions as a valid alternative feedback method in case ratings are not available, or as a feedback method 'during' content consumption instead of 'after' finishing the consumption. This answers our third research question.

Besides the emotion score, we also studied the implicit feedback score (equation 4), which is the combination of emotion and attention score. Nevertheless, the variation in the attention score was limited for our user test, since all used videos are rather short (3-5 minutes). We suspect that the duration of the videos is too short to build up intense emotional moments that make users inclined to cover their eyes or mouth. Moreover, the videos are too short to witness a decreasing level of attention (e.g., falling asleep). Therefore, we expect that the attention score and implicit feedback score might be better suited as implicit feedback for content items with a longer duration.

5 Discussion

During the user test, it became clear that people do not express their emotions much during video watching, even not if the videos contain scenes with intense emotions as selected in our test. This was visible in the results of the automatic emotion recognition (Section 4.5) as well as in the questionnaire (Section 4.2) that assesses users' emotion expression during video watching. Although the videos of the experiment were carefully chosen to evoke emotions, only 61% of the users stated that they expressed emotions during video watching.

Happiness is expressed most clearly, and is best recognized. It is the only emotion that reached the maximum probability value of 1.0, e.g., for person 13 as visible in Figure 9. For the other basic emotions, the recognition services typically register probabilities that are much lower. The second most recognizable emotion was sadness. It has a maximum value over all users of 0.68, with only 15% of the test users scoring a sadness value of 0.60 or higher (for the sad video). In the questionnaire, most users (61%) confirmed that they express their sadness while watching a video that evokes this emotion. For fear, the maximum registered value over all test users was only 0.27 (during the fearful video in the online test). Fear is the most difficult emotion to recognize, as was also discussed in the offline test.

For this experiment, the emotion fingerprint was constructed by aggregating the emotion values of all users. A big challenge is to identify the correct expected emotions and their probability values for the fingerprint spectrum. For this, we propose the following guidelines: 1) Limit the fingerprint to a few emotions that are clearly expressed in the video scene. 2) Some emotions, such as fear, are more difficult to detect than others, such as happiness. The emotion probabilities from the facial recognition services are often much lower for the difficult emotions. This should be reflected in the values of the fingerprint. 3) Limit the comparison of expected and expressed emotions to the key scenes of the movie. Recognized emotions during scenes without emotions might be due to other causes than the video.

After the experiment, many users stated that they liked the approach of emotion recognition, attention measurements, and behavior analysis for automatic feedback and improving their recommendations. However, some users were a bit concerned that photos were taken of them during video watching. They stated that this is no problem for short videos during the experiment. But 6 of the 17 users who answered the last question of the survey, do not want to use these facial recognition approaches for all their video watching activities. The main concerns are privacy issues and the feeling of being watched. Therefore, this approach is more suitable for mobile device, such as smartphones, which are also more personal devices than a smart TV for example. Using mobile devices, users are also more in control of what is visible for the front-facing camera, in contrast to the camera of a smart TV in the living room.

6 Conclusion

An Android app was developed to investigate if facial recognition services can be used as a tool for automatic authentication, user profiling, and feedback gathering during video watching. The goal is to use these profiles and feedback as input for a recommender system. In contrast to ratings, the feedback based on facial recognition is available during content playback. An evaluation with a test panel showed that the authentication is almost perfect. The estimation of gender and age was in most cases accurate enough to cope with the cold-start problem by recommending movies typical for the user's age and gender. Facial analysis can be used to derive automatic feedback from the user during video watching. Closed eyes, looking away (head pose, attention level), covering eyes or mouth (occlusion), etc., are typical indications that the user does not want to see the video, and can be considered as negative implicit feedback for the recommender. During video watching, emotions expressed by the user's face were recognized. Happiness could be most accurately detected. By comparing the recognized emotions with an emotion fingerprint, we calculated a user feedback value, which is positively correlated to the user's star rating. This confirmed the assumption that recognized emotions can be considered as valuable implicit feedback for the recommender. This way, our application achieves the objective of a more user-friendly user-recommender interaction. Through a questionnaire, users confirmed that their preferences match the recommendations based on the gathered feedback for previously watched videos. Taking photos or making videos with the front-facing camera has been expressed as a privacy-sensitive aspect by our test users and will be further tackled in future research.

References

1. Adomavicius, G., Sankaranarayanan, R., Sen, S., Tuzhilin, A.: Incorporating contextual information in recommender systems using a multidimensional approach. ACM Transactions on Information Systems (TOIS) **23**(1), 103–145 (2005)
2. Arapakis, I., Moshfeghi, Y., Joho, H., Ren, R., Hannah, D., Jose, J.M.: Enriching user profiling with affective features for the improvement of a multimodal recommender system. In: Proceedings of the ACM International Conference on Image and Video Retrieval, pp. 1–8 (2009)
3. Arapakis, I., Moshfeghi, Y., Joho, H., Ren, R., Hannah, D., Jose, J.M.: Integrating facial expressions into user profiling for the improvement of a multimodal recommender system. In: 2009 IEEE International Conference on Multimedia and Expo, pp. 1440–1443. IEEE (2009)
4. Baltrunas, L., Makcinskas, T., Ricci, F.: Group recommendations with rank aggregation and collaborative filtering. In: Proceedings of the fourth ACM conference on Recommender systems, RecSys '10, pp. 119–126. ACM, New York, NY, USA (2010)
5. Bi, H., Li, N., Guan, H., Lu, D., Yang, L.: A multi-scale conditional generative adversarial network for face sketch synthesis. In: 2019 IEEE International Conference on Image Processing (ICIP), pp. 3876–3880. IEEE (2019)
6. Bimbot, F., Bonastre, J.F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-García, J., Petrovska-Delacrétaz, D., Reynolds, D.A.: A tutorial on text-independent speaker verification. EURASIP Journal on Advances in Signal Processing **2004**(4), 101962 (2004)
7. Buthpitiya, S., Zhang, Y., Dey, A.K., Griss, M.: N-gram geo-trace modeling. In: International Conference on Pervasive Computing, pp. 97–114. Springer (2011)
8. Chauhan, M., Sakle, M.: Study & analysis of different face detection techniques. International Journal of Computer Science and Information Technologies **5**(2), 1615–1618 (2014)
9. Cho, M., Kim, T., Kim, I.J., Lee, S.: Relational deep feature learning for heterogeneous face recognition. arXiv preprint arXiv:2003.00697 (2020)
10. CommonSenseMedia: You know your kids. we know media and tech. together we can build a digital world where our kids can thrive. (2019). Available at `https://www.commonsensemedia.org/about-us/our-mission`
11. De Pessemier, T., Dooms, S., Martens, L.: Comparison of group recommendation algorithms. Multimedia Tools and Applications **72**(3), 2497–2541 (2014)
12. De Pessemier, T., Verlee, D., Martens, L.: Enhancing recommender systems for tv by face recognition. In: 12th International Conference on Web Information Systems and Technologies (WEBIST 2016), vol. 2, pp. 243–250 (2016)

13. Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. Journal of personality and social psychology **17**(2), 124 (1971)
14. Ekstrand, M.D.: The lkpy package for recommender systems experiments. Computer Science Faculty Publications and Presentations 147, Boise State University (2018). DOI 10.18122/cs_facpubs/147/boisestate. URL `https://md.ekstrandom.net/pubs/lkpy`
15. Face++: Cognitive services - leading facial recognition technology (2019). Available at `https://www.faceplusplus.com/`
16. Fan, D.P., Zhang, S., Wu, Y.H., Liu, Y., Cheng, M.M., Ren, B., Rosin, P.L., Ji, R.: Scoot: A perceptual metric for facial sketches. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5612–5622 (2019)
17. Feng, T., Yang, J., Yan, Z., Tapia, E.M., Shi, W.: Tips: Context-aware implicit user identification using touch screen in uncontrolled environments. In: Proceedings of the 15th Workshop on Mobile Computing Systems and Applications, pp. 1–6 (2014)
18. Hassan, M.M., Alam, M.G.R., Uddin, M.Z., Huda, S., Almogren, A., Fortino, G.: Human emotion recognition using deep belief network architecture. Information Fusion **51**, 10–18 (2019)
19. Hossain, M.S., Muhammad, G.: Emotion recognition using deep learning approach from audio–visual emotional big data. Information Fusion **49**, 69–78 (2019)
20. Huang, Y., Wang, Y., Tai, Y., Liu, X., Shen, P., Li, S., Li, J., Huang, F.: Curricularface: Adaptive curriculum learning loss for deep face recognition. arXiv preprint arXiv:2004.00288 (2020)
21. IMDb: Ratings and reviews for new movies and tv shows. (2019). Available at `https://www.imdb.com/`
22. Jain, A.K., Bolle, R., Pankanti, S.: Biometrics: personal identification in networked society, vol. 479. Springer Science & Business Media (2006)
23. Joho, H., Jose, J.M., Valenti, R., Sebe, N.: Exploiting facial expressions for affective video summarisation. In: Proceedings of the ACM international conference on image and video retrieval, p. 31. ACM (2009)
24. Kairos: Serving businesses with face recognition (2019). Available at `https://www.kairos.com/`
25. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), pp. 46–53. IEEE (2000)
26. Li, Y., Hu, H., Zhou, G.: Using data augmentation in continuous authentication on smartphones. IEEE Internet of Things Journal **6**(1), 628–640 (2018)
27. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 94–101. IEEE (2010)
28. Masthoff, J.: Group recommender systems: Combining individual models. In: Recommender systems handbook, pp. 677–702. Springer (2011)
29. Microsoft-Azure: Face api - facial recognition software (2019). Available at `https://azure.microsoft.com/en-us/services/cognitive-services/face/`
30. Nickel, C., Wirtl, T., Busch, C.: Authentication of smartphone users based on the way they walk using k-nn algorithm. In: 2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 16–20. IEEE (2012)
31. Noldus: Facereader - facial expression recognition software (2019). Available at `https://www.noldus.com/human-behavior-research/products/facereader`
32. Qi, M., Lu, Y., Li, J., Li, X., Kong, J.: User-specific iris authentication based on feature selection. In: 2008 International Conference on Computer Science and Software Engineering, vol. 1, pp. 1040–1043. IEEE (2008)
33. Schein, A.I., Popescul, A., Ungar, L.H., Pennock, D.M.: Methods and metrics for cold-start recommendations. In: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '02, pp. 253–260. ACM, New York, NY, USA (2002)
34. Shabrina, N., Isshiki, T., Kunieda, H.: Fingerprint authentication on touch sensor using phase-only correlation method. In: 2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), pp. 85–89. IEEE (2016)

35. Soleymani, M., Pantic, M., Pun, T.: Multimodal emotion recognition in response to videos. IEEE transactions on affective computing **3**(2), 211–223 (2011)
36. Tkalčič, M., Košir, A., Tasič, J.: Affective recommender systems: the role of emotions in recommender systems. In: Joint proceedings of the RecSys 2011 Workshop on Human Decision Making in Recommender Systems (Decisions@RecSys'11) and User-Centric Evaluation of Recommender Systems and Their Interfaces-2 (UCERSTI 2) affiliated with the 5th ACM Conference on Recommender, pp. 9–13 (2011)
37. Wang, J., Zhang, J., Luo, C., Chen, F.: Joint head pose and facial landmark regression from depth images. Computational Visual Media **3**(3), 229–241 (2017)
38. Yang, M.H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: A survey. IEEE Transactions on pattern analysis and machine intelligence **24**(1), 34–58 (2002)
39. Yu, Z., Zhou, X., Hao, Y., Gu, J.: Tv program recommendation for multiple viewers based on user profile merging. User Modeling and User-Adapted Interaction **16**, 63–82 (2006)