Enhancing Recommender Systems for TV by Face Recognition

Toon De Pessemier, Damien Verlee and Luc Martens
iMinds, Ghent University, Technologiepark 15, B-9052 Ghent, Belgium

Keywords: Recommender System, Face Recognition, Face Detection, TV, Emotion Detection.

Abstract: Recommender systems have proven their usefulness as a tool to cope with the information overload problem for many online services offering movies, books, or music. Recommender systems rely on identifying individual users and deducing their preferences from the feedback they provide on the content. To automate this user identification and feedback process for TV applications, we propose a solution based on face detection and recognition services. These services output useful information such as an estimation of the age, the gender, and the mood of the person. Demographic characteristics (age and gender) are used to classify the user and cope with the cold start problem. Detected smiles and emotions are used as an automatic feedback mechanism during content consumption. Accurate results are obtained in case of a frontal view of the face. Head poses deviating from a frontal view and suboptimal illumination conditions may hinder face detection and recognition, especially if parts of the face, such as eyes or mouth are not sufficiently visible.

1 INTRODUCTION

Recommender systems are software tools and techniques providing suggestions for items to be of interest to a user (Resnick and Varian, 1997). By filtering the content and selecting the most appropriate items according to the user's personal preferences, recommender systems can help to overcome the problem of information overload. The suggestions provided are aimed at supporting their users in various decision-making processes, such as what items to buy, what movies to watch, or what news to read.

For years, traditional recommender systems are very successful for desktop internet applications. With the growing popularity of mobile devices and their increased connectivity, recommender systems have expanded their area of application to the mobile platform. Additional information about the user, which is accessible through the camera, microphone, or sensors such as gyroscope and GPS, allow to further improve the accuracy of the recommendations and adjust them to the current user context (De Pessemier et al., 2014b). The accessibility and popularity of operating systems such as Android further stimulate the development of recommendation tools. A similar evolution can be expected for the television platform, on which smart TVs (running Android) are becoming more popular. Recommender systems for smart TVs can assist users in selecting the TV content that matches the user's preferences best.

For the television platform, additional challenges for recommender systems emerge, such as the limited interaction possibilities (a TV viewer is "leaning back" in the sofa, using only a remote control as a means of interaction), the undesirability of user accounts (the television is a shared devices on which users are not used to log in), and the consumption of content in group (people often watch television together). To cope with the problem of limited user feedback and user identification, we propose a recommender system that uses face detection and recognition. An overview of existing research related to the face recognition problem is provided in Section 2. Section 3 explains how and which face detection and recognition services have been used in our recommender system. Section 4 elaborates on the advantages of face recognition for recommender systems. The cold start problem can be alleviated by recognizing demographic characteristics of the user such as age and gender. Recognizing emotions of users during content watching can be used to derive implicit feedback for the content. For group recommendations, identifying the people in front of the TV can be automated by using face recognition. In Section 5, the used face recognition services are evaluated. Details are provided about the accuracy of estimating peoples age, recognizing gender, and detecting emotions from a picture of a person's face. Furthermore, the accuracy of the face detection process is evaluated for pictures with various poses of the head and with different

Pessemier, T., Verlee, D. and Martens, L.
Enhancing Recommender Systems for TV by Face Recognition.
In *Proceedings of the 12th International Conference on Web Information Systems and Technologies (WEBIST 2016) - Volume 2*, pages 243-250
ISBN: 978-989-758-186-1

illumination conditions. Finally, Section 6 draws conclusions and points to future work.

2 RELATED WORK

Various solutions for tracking people and face recognition have been proposed in literature. The Reading People Tracker (Siebel, 2015) is an open source software tool for tracking people in camera images for visual surveillance purposes. It is often used for automatic visual surveillance systems for crime detection and prevention. The software is written in C++ and therefore difficult to integrate into an Android application running on a smart TV.

TrackLab is a tool developed by Noldus for recognition and analysis of spatial behavior (Noldus, 2015b). TrackLab facilitates the development of interactive systems that respond in real-time to the location or spatial behavior of subjects being tracked. The collected data can be visualized by showing tracks on a map or by heat maps of the aggregated location data of multiple people. Statistics can provide insights into the current position of a user, and when a user enters or leaves a specific room (e.g. the TV room). Since TrackLab is a software tool running on Windows with minimum requirements of 1 GB hard disk space, 1 GB RAM, and a 1Ghz CPU, it is less suitable for integration into an Android Smart TV.

FaceR is a commercial service developed by Animetrics (Animetrics, 2015) that offers a REST API (Application Programming Interface) for face recognition based on pictures. The FaceR service is able to provide the coordinates, orientation, and pose of the detected face. Compute-intensive tasks, such as image analysis or face template generation, are handled by the server layer, a set of clonable servers to ensure scalability. The core business of this service is identity management and authentication with use cases in law enforcement and commercial and consumer markets.

The FaceReader service is the facial expression analysis tool of Noldus (Noldus, 2015a). FaceReader can detect the position of mouth, eyes, and eyebrows. This service can analyze six basic facial expressions (emotions) and detect the gaze direction, head orientation, and person characteristics (gender, race, age, wearing glasses, etc.). An API is available to serve as an interface between FaceReader and different software programs using it, thereby facilitating the integration of the service.

Another service for emotion recognition is EmoVu (Eyeris, 2015). Their deep learning based technology enables to recognize emotions from facial micro-expressions. The recognition process can handle pictures as well as videos as input. The service outputs the coordinates and orientation of the face, gender, an age category, and recognized emotions with an intensity score.

Rekognition is a service that can recognize more than gender, age, and emotions (Orbeus, 2015). Rekognition has the ambition of recognizing concepts, such as a party, the beach, a cat, the Golden Gate Bridge, etc., from pictures as well. Although the concept recognition is still under development, the beta version of Rekognition can return the best five guesses with confidence scores.

A few studies have combined face detection and recognition with recommender systems. Recognized emotions can be used to automatically derive feedback for the content. This way, the topical relevance of a recommended video has been predicted by analyzing affective aspects of user behavior (Arapakis et al., 2009). In addition, affective aspects can be used to model user preferences. The underlying assumption is that affective aspects are more closely related to the user's experience than generic metadata, such as genre (Tkalcic et al., 2010). The end goal is to incorporate emotions into the recommendation process as well. Recommending music according to the user's emotion is a typical use case (Kuo et al., 2005). However, in this model for emotion-based music recommendation, the user's emotions are obtained by querying them in stead of by automatic recognition.

3 FACE DETECTION AND RECOGNITION

In this section, we describe the face detection and recognition process of our recommender system for Android smart TVs. Face detection and recognition is performed in two subsequent phases.

In phase 1, human faces are detected using the face detection mechanism of the Android API (Meier, 2015) (Android API level 14). Android's face detection mechanism is commonly used in applications that use the camera to focus on people's face. It can detect up to sixteen faces simultaneously. Figure 1 shows a screenshot of the application with rectangles indicating the detected faces in an image originating form the camera. These rectangles are only shown in the test version of the developed application. In the final version of the application, face detection is performed as a background process, without bothering the end-users. A face detection listener is coupled to the Android camera object to check continuously who is in front of the TV. If one or more faces are detected by the Android system, the listener is notified and a picture is taken. The camera focus is automatically adjusted to the detected faces.

In phase 2, the taken pictures are used as input for the face recognition process. In our implementation we use two different face recognition services: Face++ and SkyBiometry. Face++ (Face++, 2015) is a real-time face detection and recognition service. The results of the Face++ recognition process are the positions of the faces with detailed X,Y coordinates for the eyes, nose, and mouth. Besides, the face recognition service can detect glasses. More important for our recommender system is the service's estimation of the person's age, gender, and race, together with a confidence value for each attribute. Another interesting outcome of the service is the degree to which the subject smiles with an associated confidence value. Face++ stores the results of the face recognition processes in a database to compare future recognition requests. If a new face recognition request shows similarities with a previously recognized face, a similarity indicator is specifying the resemblance. If this resulting similarity indicator is above a certain threshold, our application assumes that this person is a returning user and therefore already registered in the system.

SkyBiometry (SkyBiometry, 2015) is a service very similar to Face++ but uses a different computer vision algorithm. It is a cloud based face detection and recognition service that is available through an API. The service is able to detect multiple faces at different angles in a picture and also provides the location of the eyes, nose, and lips. The service makes an assessment of the presence of glasses (dark glasses or not), the fact that the person is smiling and the lips are sealed or open, whether the person's eyes are open or not, the person's gender, and the person's mood (e.g., happy, sad, angry, surprised, disgusted, scared, neutral). For each of these attributes, a percentage is indicating the confidence value of the estimation. The age of a person is specified by a point estimator. Faces already known by the service can be recognized.

For an optimal face detection and recognition, a picture is taken using the camera of the smart TV and sent for analysis to these two services every fifteen seconds. The big advantage of using two face detection and recognition services, using different algorithms, is the increased accuracy by combining them. In case the two services do agree, the results can be used with a high degree of certainty. If they do not agree, one of them is chosen (Section 5) or a new picture is send for reanalysis.

This way our application enables automatic authentication of users in front of the TV. To provide users feedback on this authentication process, the recognized persons are shown in the user interface. Therefore, the captured picture is cropped, so that only the head is remaining, and used as profile picture in the application (Figure 1, left side).

4 RECOMMENDER SYSTEM

4.1 Cold Start Solution

Traditional recommender systems suffer from the new user problem, i.e. the issue that recommender systems cannot generate accurate recommendations for new users who have not yet specified any preference. To cope with the new user problem (also known as the *cold start problem*), our system recommends videos for new users based on the derived demographic characteristics of the user, such as age and gender. These user characteristics are matched to the demographic breakdowns of the ratings for movies on IMDb.com.

Figure 2 shows an example of such a demographic breakdown for the ratings of the movie "The Twilight Saga: Breaking Dawn - Part 1". For this movie, a significant difference in rating behavior of 1.8 stars is visible for men and women. For specific age groups, these differences may vary. For example, a difference of 2.3 stars is witnessed for people under 18, whereas the age group of 45+ has a difference of 0.9 between men and women. The ratings of the specific age group and gender are selected based on the user's gender and age as estimated by the face recognition service. Subsequently, the user's preference for a movie is predicted based on these ratings. As soon as more detailed preferences of the user become available (e.g. through ratings), these are taken into account by using a standard collaborative filtering system. These collaborating filtering (CF) recommendations are combined with the recommendations based on demographics (demo) using a weighted average.

$$Rec_{combined} = w_{CF} \cdot Rec_{CF} + w_{demo} \cdot Rec_{demo} \quad (1)$$

As more rating data of the user becomes available, the collaborative filter is expected to become more accurate and therefore the weight of the collaborative filter (w_{CF}) is increasing while the weight of the demographics (w_{demo}) is decreasing. In Figure 1, these recommendations are visualized on the right side of the screen by means of the posters of the movies. Posters and metadata of movies are retrieved using the TMDb API (Themoviedb.org, 2015).

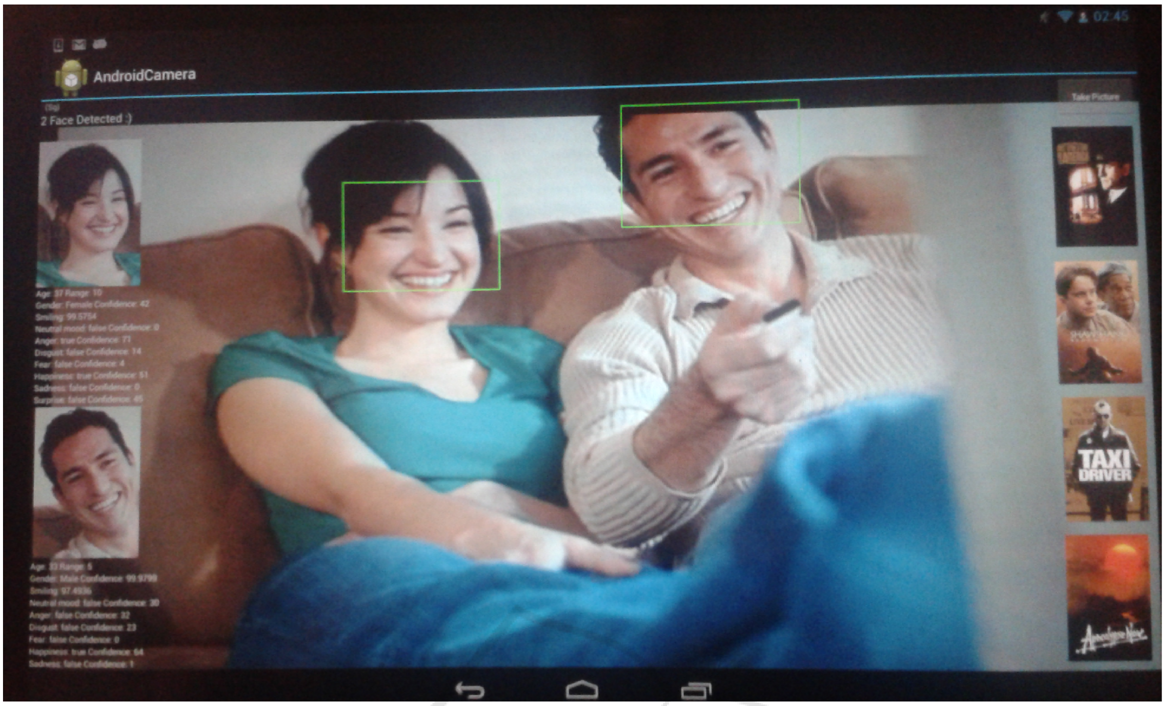

Figure 1: A Screenshot of application illustrating the face detection mechanism by means of squares around the faces.

4.2 Implicit Feedback by Detecting Emotions

Collecting these rating data is another issue of recommender systems for TV. Because of the passive attitude of the typical TV viewer, evaluating the content by specifying a rating is often skipped. Another difficulty is the timing of the ratings. Whereas feedback during video watching can be useful for suggesting alternative TV content, ratings are typically collected when content playback is finished.

Therefore, our system *automatically collects feedback* during content playback based on the detected emotions of the viewers. Emotion detection is handled by the used face recognition service (SkyBiometry, 2015) that keeps the complexity of this process in the cloud. The extent to which users are engaged in a TV show can be used to estimate their interests in the show. So, expressed emotions are considered as user engagement and used as a feedback mechanism. The stronger the detected emotions, the stronger the user's engagement, and the stronger the feedback signal.

Since not all emotions expressed by the viewer are provoked by the TV content, the user's detected emotions are compared to the emotions that can be expected from the content. Therefore, a database is created with typical emotions for different sections of each content item. These typical emotions are calculated by aggregating the recognized emotions of all users who watched the content item in the past. E.g., for a comical scene, 'happy' turns out to be the predominant emotion for most viewers, which is subsequently considered as the typical emotion for this scene. Some viewers may also be 'surprised' by the scene, which can be considered as another emotion evoked by the scene. If the user's expressed emotions are similar to the typical emotions as expressed by the community, then these emotions are considered as provoked by the content. Each emotion i detected during scene watching is associated with a weight w_i. The value of this weight is determined by the extend to which this emotion is recognized (output face recognition) and the number of people who expressed this emotion during the scene.

The extent to which users show their emotions depends on the character of the person (extrovert or introvert). Some people clearly show their emotions while others stay rather neutral all the time. Therefore, the confidence values of the face recognition services regarding the detected emotions are normalized for each individual user. This normalization is based on all detected emotions of an individual user over different content items. In the current implementation, the average confidence value over all emotions of the individual is used to normalize the intensity (confidence value) of a detected emotion of that user. Each of these normalized confidence values express user engagement in terms of a different emotion

e_i. Subsequently the recognized emotions are multiplied by the weights of the typically expected emotions for the scene. This overall value of user engagement is considered as an implicit feedback value, $Feedback_{emo}$.

$$Feedback_{emo} = \frac{\sum_{i=0}^{n} e_i \cdot w_i}{n} \tag{2}$$

4.3 Group Recommendations

Although most recommender systems are designed to serve individual users, many activities, such as watching TV, are often group activities performed by multiple friends or family members together. If multiple people are detected in the room watching television together, the recommendations cannot be limited to the preferences of one individual, but group recommendations have to be generated.

One of the problems that comes with group recommenders is identifying the group for which recommendations have to be generated. In many cases, only one person is controlling the system (browsing content, making a selection, etc.), i.e. the person who has the remote control in case of TV watching. A traditional solution to express the group's members is asking the users to log on with their personal account. However, on television sets people are not used to log on and specifying passwords with a remote control is a devious activity. Moreover, if a person is leaving the room, this user should log off to specify the change in group composition.

Therefore, our system has an *automatic identification* process of the users sitting in front of the TV. The similarity indicator of the face recognition process is used to decide if a person already utilized the recommender system in the past. This way, preferences deduced from past watching behavior can be coupled to the recognized user.

These group recommendations are suggestions for the content items that are most suitable for the group as a whole. Different solutions to aggregate the preferences or recommendations of individuals and find the best items have been proposed (De Pessemier et al., 2014a). In our implementation, the *average without misery* strategy is used. The idea of this strategy is to find the optimal decision for the group, without making some group members really unhappy with this decision.

This strategy calculates $P_{Group,i}$, the prediction of a group's interests for an item i, as follows. The interests of each individual member for an item can be calculated by any recommendation algorithm. In our implementation a traditional collaborative filtering solution is used (Resnick and Varian, 1997). If the interests of one of the group members for i is predicted to be below a certain threshold θ, the item i gets a penalty and is excluded from the group recommendations ($P_{Group,i} = 0$). This prevents that the group receives recommendations for items that will make one of the group members unhappy. If the interests of all group members for item i are predicted to be above the threshold, the prediction of the group's interests for i is calculated as the average of the prediction of each member's interests, $P_{u,i}$.

$$P_{Group,i} = \begin{cases} 0, & \text{if } \exists\, u \in Group: \; P_{u,i} < \theta \\ \frac{\sum_{u \in Group} P_{u,i}}{|Group|}, & \text{otherwise} \end{cases} \tag{3}$$

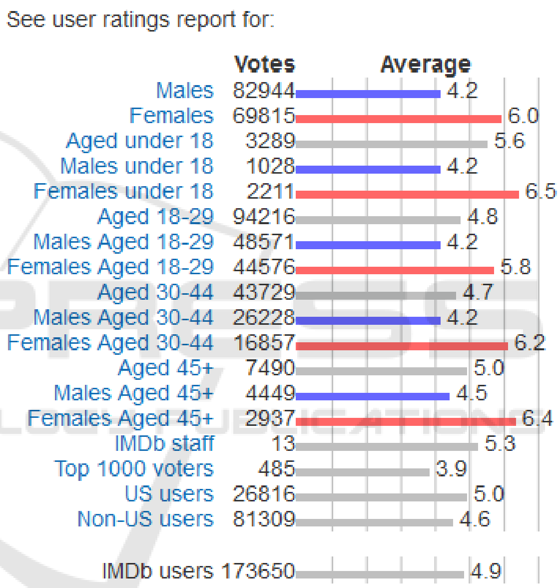

Figure 2: Demographic breakdown for the ratings of a movie on IMDb according to the age and the gender.

5 EVALUATION

5.1 Gender and Age Estimation

To evaluate the estimation of gender and age as made by the face recognition services, we used the database of Minear and Park with 180 photos of people (Park Aging Mind Lab, University of Texas, 2015). These photos, published in grey-scale, are all frontal views of people of different ages, gender, and race. All people have a neutral face expression on the photos. Half of the people are young adults (18 to 49 years old), half of them are older adults (50 to 94 years old). About half of them are men, half of them are women.

Also within the various ethnic variations, the share is about 50%-50% for men and women. For each of these photos, the exact age and gender of the person is available as ground truth. Since this database contains no photos of people under 18 years old, 34 extra photos originating from Google images were added to fill this void.

These photos are used to evaluate the accuracy of the estimation of people's age. The actual age of the person is compared to the age range that is provided by the Face++ recognition service. For each photo, the error is calculated as the difference between the predicted age range and the actual age. The age range of the prediction is always between 5 and 10 years. An average error of 2.68 years is obtained for 212 photos. (The face recognition service was unable to predict the age in two photos.)

Figure 3 shows the average prediction error per gender and age category. The most accurate results are obtained for men between 30 and 60 years old, with an error between estimated age range and actual age that is below one. Also for people below 30 years old, an accurate age estimation is obtained. In contrast to men, a lower accuracy is obtained for women between 30 and 60 years old. This difference can be explained by women who try to mask their age by wearing make-up. Age estimation is the most difficult for people above 60 years old. For these people, (small) age differences are less visible in the face.

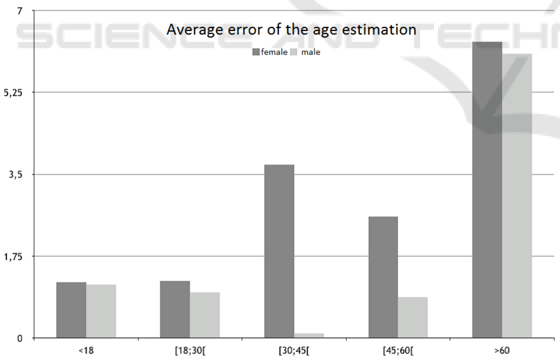

Figure 3: The accuracy of the age estimation evaluated based on the average difference (in years) between the actual age and the predicted age range, per gender and age category.

5.2 Gender Recognition

The database of Minear and Park is also used to evaluate the gender recognition of the services. Both face recognition services are combined with the aim of obtaining a higher classification accuracy. If Face++ and SkyBiometry agree and recognize the same gender, there is a high probability that this is correct. Only in a few cases in which Face++ and SkyBiometry agree, but both have a low confidence value, a wrong recognition was made. If Face++ and SkyBiometry disagree about the gender for a person above 18 years old, the recognition of SkyBiometry is used since this service obtained the most accurate results. For people under 18 years old, the recognition of Face++ is the most accurate and therefore used as final recognition.

Table 1 shows the results of the gender recognition by combining Face++ and SkyBiometry. An overall accuracy of 92.52% is obtained for the combination of services, whereas the individual services obtained an accuracy of 75.23% (Face++) and 83.64% (SkyBiometry).

Table 1: Accuracy of the gender recognition.

	Males	Females
Recognized as male	98	6
Recognized as female	10	100

5.3 Emotion Recognition

For recognizing emotions in people's faces, the SkyBiometry service is used. The output of Face++ contains a value indicating the extent to which a person is smiling, but recognizes no specific emotion, such as surprised, disgusted, or scared. To evaluate the emotion recognition, the Cohn-Kanade AU-Coded Expression Database is used (Affect Analysis Group - Research Lab at the University of Pittsburgh, 2015). This is a collection of photos of persons' faces with an emotional value (happy, sad, angry, surprised, disgusted, scared or contempt) (Lucey et al., 2010). The database contains no photos with a neutral face expression. Photos with the emotion 'contempt' are ignored since this emotion cannot be recognized by SkyBiometry.

This database of photos is used to evaluate the emotion recognition of the SkyBiometry service. For each photo, the emotion as stated in the database is compared to the recognized emotion. An overall accuracy of 83.88% is obtained. Figure 4 shows the obtained accuracy of the emotion recognition per emotion. Happy, disgusted, and scared are emotions which are relatively easy to recognize in contrast to sad, surprised, and angry.

5.4 Head Orientation

In ideal conditions, a front facing picture of the TV viewer is available as input for the recognition process. However in realistic scenarios (e.g. lying on the sofa watching TV), the TV viewer's head can be observed under various orientations. Head poses deviating from a frontal view can introduce difficulties for

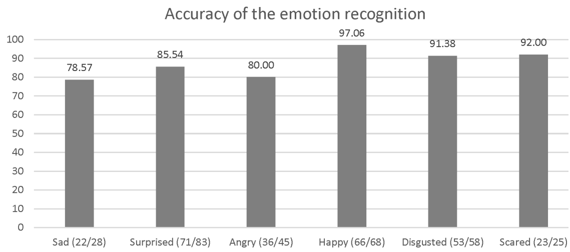

Figure 4: The accuracy of emotion recognition.

the face detection process. Therefore, we evaluated the influence of various head orientations on the accuracy of the face detection mechanisms of Face++ and SkyBiometry by using photos of the Head Pose Image Database (Gourier et al., 2015a). The Head Pose Image Database contains 2790 face images of 15 persons with variations of pan and tilt angles from -90 to +90 degrees. For every person, 93 images (93 different poses) are available. For each of these poses, the face detection of the Face++ and SkyBiometry service is tested.

Figure 5 shows the results for the Face++ service for the various poses of the head. Blue dots stand for detected faces. Red dots correspond to undetected faces. In Figure 6, the results obtained with SkyBiometry are shown. Comparison of the two services shows that Face++ is more sensitive to head poses deviating from a frontal view. Face++ is able to detect the face for 24.73% of the different head poses, whereas SkyBiometry can successfully detect the face for 66.67% of the poses.

Face detection can only be successful if a considerable part of the face is visible. The blue dots in Figure 5 and 6 have the shape of a funnel: wide at the top and narrow at the bottom. This means that for variations with tilt angles, faces leaning backward can better be detected than faces leaning forward.

5.5 Illumination Conditions

Another important factor influencing the accuracy of face detection and recognition services is the illumination condition. To evaluate this influence, the photos of the Yale Face Database were used (Gourier et al., 2015b). This database contains 5760 single light source images of 10 people, each seen under 576 viewing conditions (9 poses x 64 illumination conditions). The position of the illumination is denoted by the azimuth and elevation of the single light source direction.

Overall, Face++ is able to detect the face in 76.56% of the pictures whereas SkyBiometry has a successful detection for 70.31% of the various illumination conditions. Undetectable faces are mainly due to insufficient exposure of the face: a face that is completely in the shade, eyes and mouth that are not sufficiently visible, etc. These cases generally correspond to an illumination with an azimuth outside the range $[-35^\circ, +35^\circ]$ or an elevation outside the range $[-50^\circ, +50^\circ]$ (measured from a frontal view). Using such an illumination, a considerable part of the face is not exposed by the light.

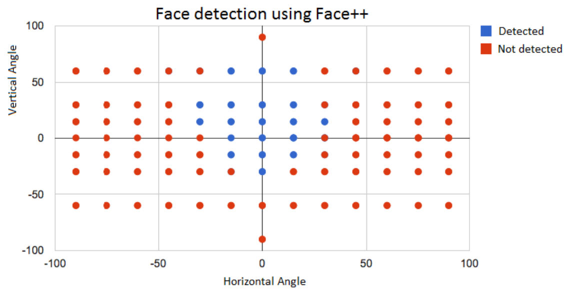

Figure 5: The accuracy of the face detection of Face++ under different orientations of the head.

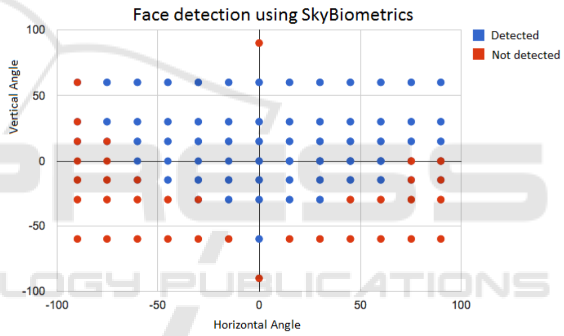

Figure 6: The accuracy of the face detection of SkyBiometry under different orientations of the head.

6 CONCLUSIONS

Smart TVs equipped with a camera and microphone allow to develop additional features for enhancing the TV watching experience. Face recognition services can be used for improving the user experience of recommender systems by automatic feedback generation and user identification. Using cloud-based face recognition services reduces the computational requirements of TV sets and allow a continuous improvement of the accuracy. Two of these face recognition services, Face++ and SkyBiometry are investigated in detail. These services proved to be a valuable tool to estimate the age and recognize the gender of people based on a photo of their face. By combining the results of both services, we improved the accuracy of the recognition process. Furthermore, face recognition services can be used to recognize emotions from people's face. Automatic feedback for the content can be generated by matching the recognized emotions to the emotions that are expected to be evoked by the content. In realistic environments, such as a living room, face recognition has to cope with some serious difficulties such as head poses deviating from a frontal view and suboptimal illumination conditions. If important parts of the face, such as eyes or mouth, are not sufficiently visible, face detection and recognition fail. An interesting research track for future work is to evaluate the accuracy of the automatically generated feedback based on the recognized emotions by a test panel. Also privacy and data security aspects should be investigated, especially if personal information is sent to external services. Possible solutions can include data encryption mechanisms or privacy-preserving architectures that limit the amount of disclosed data of the TV viewers.

REFERENCES

Affect Analysis Group - Research Lab at the University of Pittsburgh (2015). Cohn-kanade au-coded expression database. Available at http://www.pitt.edu/∼emotion/ck-spread.htm.

Animetrics (2015). Facer identity management system: Connecting mobile devices with photographic clouds. Available at http://animetrics.com/cloud-web-service-architecture/.

Arapakis, I., Moshfeghi, Y., Joho, H., Ren, R., Hannah, D., and Jose, J. (2009). Integrating facial expressions into user profiling for the improvement of a multimodal recommender system. In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, pages 1440–1443.

De Pessemier, T., Dooms, S., and Martens, L. (2014a). Comparison of group recommendation algorithms. *Multimedia Tools and Applications*, 72(3):2497–2541.

De Pessemier, T., Dooms, S., and Martens, L. (2014b). Context-aware recommendations through context and activity recognition in a mobile environment. *Multimedia Tools and Applications*, 72(3):2925–2948.

Eyeris (2015). Emovu emotion recognition software. Available at http://http://emovu.com.

Face++ (2015). Leading face recognition on cloud. Available at https://http://www.faceplusplus.com/.

Gourier, N., Hall, D., and Crowley, J. L. (2015a). Head pose image database. Available at http://www-prima.inrialpes.fr/perso/Gourier/Faces/HPDatabase.html.

Gourier, N., Hall, D., and Crowley, J. L. (2015b). The yale face database. Available at http://vision.ucsd.edu/content/yale-face-database.

Kuo, F.-F., Chiang, M.-F., Shan, M.-K., and Lee, S.-Y. (2005). Emotion-based music recommendation by association discovery from film music. In *Proceedings of the 13th Annual ACM International Conference on Multimedia*, MULTIMEDIA '05, pages 507–510, New York, NY, USA. ACM.

Lucey, P., Cohn, J., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). 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*, pages 94–101.

Meier, R. (2015). Android developers blog. Available at http://android-developers.blogspot.be/2015/08/face-detection-in-google-play-services.html.

Noldus (2015a). Facereader: Facial expression recognition software. Available at http://www.noldus.com/facereader.

Noldus (2015b). Tracklab. Available at http://www.noldus.com/innovationworks/products/tracklab.

Orbeus (2015). Rekognition - integrated visual recognition solution based in the cloud. Available at https://rekognition.com/.

Park Aging Mind Lab, University of Texas (2015). Face database. Available at http://agingmind.utdallas.edu/facedb/view/normed-faces-by-kristen-kennedy.

Resnick, P. and Varian, H. R. (1997). Recommender systems. *Commun. ACM*, 40(3):56–58.

Siebel, N. T. (2015). People tracking - the reading people tracker. Available at http://www.siebel-research.de/people_tracking/reading_people_tracker/.

SkyBiometry (2015). Cloud-based face detection and recognition api. Available at https://www.skybiometry.com/.

Themoviedb.org (2015). TMDb API. Available at https://www.themoviedb.org/documentation/api.

Tkalcic, M., Burnik, U., and Kosir, A. (2010). Using affective parameters in a content-based recommender system for images. *User Modeling and User-Adapted Interaction*, 20(4):279–311.