Abstract

The Looking Back Screens demonstrator presents a high end ‘multi-level’ face analysis system that observes facial characteristics, expressions, gaze and (facial) gestures of viewers and integrates these viewer observations in various interactive ‘games’ with an in-focus viewer.

1. Introduction

Since we all spend so much time nowadays looking at the screen of a computer, TV, or mobile, it seems about time these screens start to look back at us, their viewers, and learn to adapt their screen presentations to our interests. The Looking Back Screens (LBS) demonstrator presents a high end ‘multi-level’ face analysis system that observes facial characteristics, expressions, gaze and (facial) gestures of viewers and integrates these observations in various interactive modes of presentation of information.

A main application area for screens that look back at you is known as ‘narrow casting’, or ‘in store digital media’, i.e. large screens in public places such as shops and offices, that present pre-programmed content - information, advertisements and entertainment- from a dedicated server. Narrow casting has been hyped but initial experiences have sometimes been disappointing. Tesco hung its shops full with big screens, but found out after a while that -almost- nobody was watching. To make localised and dedicated content presentation effective, the presentation system has to be able to see who is watching, and then find ways to engage that public.

2. The LBS demonstrator

The demonstrator physically consists of three side by side mounted LCD screens (22” or larger) at eye level, a handful of camera’s, i.e. 5 standard webcams, almost hidden above or between screens, and a few pc’s (3). LBS continuously scans for possible viewers, and registers general characteristics of viewers found, and in particular aims to find and track a single in-focus viewer to engage in interaction. Interactions can be initiated by dialog texts, greetings, requests and invitations for further ‘games’ or playful interactions that can be answered by yes/no head gestures, or may start implicitly. Interaction games include: pose emotions, ‘Could you look angry for me? ’; mimicking, the facial expression of the viewer is imitated by an avatar; implicit profiling and recommending.

In order to manage the complexities of such a multimodal and multilevel system, the design of the system follows a highly modular “embedded systems” approach [5]. Several sub-components continuously run their individual tasks, while a decision layer grabs and combines the sub-results to form decisions and/or interactions when necessary.

3. System components

3.1. Face analysis and recognition

One of the key components of the LBS demonstrator is a face analysis system based on VicarVision’s FaceReader™ software. FaceReader is a software environment for emotional expression analysis currently used world-wide by research institutes and commercial organizations.

The face analysis system describes a face in three consecutive steps. A real-time face detection system is used to find multiple faces in a camera stream. This system contains an implementation of the Viola-Jones detection algorithm for fast initial detection and a deformable template method to create more accurate face framings. The next step of the facial analysis system creates an accurate face model using the Active Appearance method described by Cootes and Taylor [1]. This face model is then classified using a battery of neural networks for various facial features such as emotional expression, gender, age, ethnicity, facial hair and facial ornaments [3]. The demonstrator also shows how a person’s identity can be determined using face recognition techniques and how this identity can be used to shape the interaction with the people visiting the demonstrator during the conference.
3.2. Head tracking and gesture recognition

The system includes a face gesture recognition module which is used to recognize agreeing or disagreeing head gestures (yes or no head nodding or shaking). The module is based on the “KLT point tracking algorithm” described by Shi and Tomasi [4], combined with a cylindrical head model. The point tracking algorithm finds features, within the bounding box of the face, which can be tracked both in the x- and y-direction, such as corner points and checkered textures. In an initial frame these key points are projected on one half of a cylinder (which forms a rough approximation of the shape of a face) with an initial orientation. In following frames the transformation of the cylinder that best fits the translation of the feature points in this frame is estimated. From this change in orientation of the cylinder (and thus change in orientation of the face) we can derive the face gestures.

The recognition of face gestures supports interesting interaction possibilities as yes or no questions posed by the LBS can be answered through natural interaction methods, used in human–to–human communication. However, an accurate and continuous estimation of head orientation is also a primary requirement for a gaze tracking system.

3.3. Gaze Tracking

The VicarVision GazeTracker™ is an eye-tracking system designed to be used with low cost off-the-shelf equipment for viewers that may move around freely over some distance. In contrast to special purpose eye-tracking systems, our system uses standard webcams without infrared capabilities and infrared lighting. Infrared lighting has clear advantages for eye-tracking as the pupil of the eye shows distinct and precise reflections when lit with infrared light. However, standard webcams are easily available everywhere and do not require special lighting that may not work well over larger viewer distances.

These requirements pose challenging constraints on the accuracy of the system and thus we do not expect the system to be as accurate as its counterparts with specialised equipment setups. In a study we showed how the Active Appearance Model is a suitable candidate for creating accurate eye models containing both texture information and positions of key landmark points within the eyes such as the pupil centre and eye corners [2]. The fit of an AAM results in a compact representation of the eyes from which a gaze direction is derived by applying neural network classifiers.

3.4. Implicit profiling and recommending

From a face, general features of the viewer such as age, gender and ethnicity can be derived, which are predictive for the viewers’ interests. Once a viewer is offered alternative images on a screen, while the gaze of the viewer is tracked, further information on viewer interests can be derived. Implicit profiling refers to the build up of a model of viewer interests on the basis of such observations, without asking explicit questions. On the basis of such a profile, new contents for presentation can be recommended or selected that may best match viewer interest.

On a number of Dutch conferences over the past year an earlier LBS demonstration system presented the viewer a selection of movies (descriptions and screenshot) based on general characteristics. When the person looks at one of the movies, its trailer begins to play. If the person is watching the trailer for a longer period the movie is added to his or her profile. Based on this extended profile a recommender system suggests three other movies to replace the three non-playing movies. For any type of content, pictures or text, such a cycle can be repeated, creating increasingly more refined profiles.