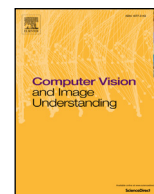




Contents lists available at ScienceDirect

## Computer Vision and Image Understanding

journal homepage: [www.elsevier.com/locate/cviu](http://www.elsevier.com/locate/cviu)

## Gender and gaze gesture recognition for human-computer interaction

Wenhao Zhang\*, Melvyn L. Smith, Lyndon N. Smith, Abdul Farooq

Centre for Machine Vision, Bristol Robotics Laboratory, University of the West of England, T Block, Frenchay Campus, Coldharbour Lane, Bristol, BS16 1QY, UK

## ARTICLE INFO

## Article history:

Received 17 April 2015

Revised 16 March 2016

Accepted 18 March 2016

Available online xxx

## Keywords:

Assistive HCI

Gender recognition

Eye centre localisation

Gaze analysis

Directed advertising

## ABSTRACT

The identification of visual cues in facial images has been widely explored in the broad area of computer vision. However theoretical analyses are often not transformed into widespread assistive Human-Computer Interaction (HCI) systems, due to factors such as inconsistent robustness, low efficiency, large computational expense or strong dependence on complex hardware. We present a novel gender recognition algorithm, a modular eye centre localisation approach and a gaze gesture recognition method, aiming to escalate the intelligence, adaptability and interactivity of HCI systems by combining demographic data (gender) and behavioural data (gaze) to enable development of a range of real-world assistive-technology applications.

The gender recognition algorithm utilises Fisher Vectors as facial features which are encoded from low-level local features in facial images. We experimented with four types of low-level features: greyscale values, Local Binary Patterns (LBP), LBP histograms and Scale Invariant Feature Transform (SIFT). The corresponding Fisher Vectors were classified using a linear Support Vector Machine. The algorithm has been tested on the FERET database, the LFW database and the FRGCv2 database, yielding 97.7%, 92.5% and 96.7% accuracy respectively.

The eye centre localisation algorithm has a modular approach, following a coarse-to-fine, global-to-regional scheme and utilising isophote and gradient features. A Selective Oriented Gradient filter has been specifically designed to detect and remove strong gradients from eyebrows, eye corners and self-shadows (which sabotage most eye centre localisation methods). The trajectories of the eye centres are then defined as gaze gestures for active HCI. The eye centre localisation algorithm has been compared with 10 other state-of-the-art algorithms with similar functionality and has outperformed them in terms of accuracy while maintaining excellent real-time performance.

The above methods have been employed for development of a data recovery system that can be employed for implementation of advanced assistive technology tools. The high accuracy, reliability and real-time performance achieved for attention monitoring, gaze gesture control and recovery of demographic data, can enable the advanced human-robot interaction that is needed for developing systems that can provide assistance with everyday actions, thereby improving the quality of life for the elderly and/or disabled.

© 2016 Elsevier Inc. All rights reserved.

## 1. Introduction

With the emergence of personal computing in the late 1970s, the concept of Human-Computer Interaction (HCI) was pushed rapidly and steadily into the environment where the individual role of science or engineering was not sufficient for addressing the urgent need for increasing the usability of computer software and operating systems (Grudin, 2011). The potential of increased accessibility to personal computers and demand for higher usability

(Karray et al., 2008) of computer platforms called for the practical need for HCI and a synthesis of science and engineering. As personal computers present more digital information to people, the way people perceive, process and respond to such data has largely altered. HCI has therefore not only become central to information science theoretically and professionally, but also has stepped into peoples' lives, offering multiple types of communication channels, i.e. modalities. These modalities gain their input via various types of sensors, mimicking human sensors including visual, audio and haptic sensors (Karray et al., 2008). They individually or combined give rise to a wide array of HCI systems. Among the various types of sensors, the visual sensor and the resulting visual based interaction modality are the most widespread, taking visual signals as

\* Corresponding author: Tel.: +44 11732 86806.

E-mail address: [wenhao.zhang@uwe.ac.uk](mailto:wenhao.zhang@uwe.ac.uk) (W. Zhang).

inputs, which include but are not limited to images/videos of faces, bodies and hands (Jaimes and Sebe, 2007). The corresponding research spans the areas of gaze tracking, gait analysis, and recognition of face, gender, age, gesture and facial expression. While face recognition reveals an individual's identity, gender/age recognition aims to gather human demographics for the system to understand human characteristics. Facial expression recognition further estimates human affection states and gathers emotional cues while gaze tracking serves to extract users' attentive information and to predict their intentions.

These modalities, independent or combined, have played different assistive roles in their corresponding HCI applications. As an example of gaze tracking, smart solutions are available that monitor the gaze direction of a driver in order to identify driver distraction/drowsiness and provide timely alerts (Tawari et al., 2014). These driver assistance systems, capable of detecting and acting on driver inattentiveness, are of great value to road safety. In the area of gesture recognition, a sterile browsing tool, 'Gestix', has been designed for doctor-computer interaction. This assistive technology provides doctors the sterility needed in an operation room where radiology images can be browsed in a contactless manner. A doctor's hand is tracked by a segmentation algorithm using colour model back-projection and motion cues from image frames (Wachs et al., 2008). Moreover, when expression recognition is concerned, an intelligent tutoring system, namely a Learning Companion, is proposed to predict when a learner might be frustrated, initiate interaction depending on the user's affective state and provide support accordingly (Kapoor et al., 2007).

With regard to the visual modality, we categorise these tasks as 'demographic recognition' (e.g. age and gender recognition) or 'behavioural recognition' (e.g. gaze analysis), according to the source and the utilisation of visual signals. In this paper, we propose a HCI strategy by combining demographic and behavioural recognition such that more natural, interactive and user-centred HCI environments can be created. More specifically, although demographic data reveal user characteristics and help define the 'initial state' and the 'general theme' of a HCI session, they cannot address the dynamic nature of a HCI session where user behaviours are constantly changing. Therefore, on the one hand, behavioural recognition reflects user attentions and intentions; on the other hand, it allows a user to issue commands and to actively interact with a HCI system through various means. As a result, the combination of demographic and behavioural recognition provides fused knowledge throughout a HCI session and better mimics a natural face-to-face interaction.

Although research in HCI has become increasingly active and sophisticated, a few issues remain unresolved that restrict most works from being transformed into assistive HCI systems that can benefit the daily life of human beings. We summarise the three major general issues that undermine the practicability of these works as follows:

- (1) Lack of accuracy in real-world scenarios. Many research works are tested on controlled databases where ideal illuminations, high-resolution images and desirable viewpoint are available. When tested under various types of scenes with dynamic environmental factors, their performance will drop severely.
- (2) Undesirable real-time performance. As powerful as they might be, sophisticated algorithms often incur large computational cost, rendering them unsuitable for real-time implementation.
- (3) High dependence on expensive or inconvenient hardware configuration. The cost and the complexity of algorithm implementation will limit the usability and applicability of any

method. Cheap yet effective methods are in high demand in order to boost assistive technologies.

In order to fill these gaps, in Section 3, we present an accurate gender recognition method that exhibits high accuracy and robustness under controlled and uncontrolled environments. This method puts Fisher Vectors at its core and encodes low-level local features (e.g. greyscale values, Local Binary Pattern (LBP), LBP histograms and the Scale Invariant Feature Transform (SIFT)) into more discriminative features for gender classification. State-of-the-art accuracy is achieved by only a linear Support Vector Machine (SVM), which further confirms the superiority of Fisher Vectors.

In Section 4, we propose a modular eye centre localisation method that makes use of isophote and gradient features extracted by its two modules. The first module performs an initial estimation of eye centre locations using isophote features from face images and filters eye centre candidates for the second module. The second module then updates the eye centre locations using only local gradient features. This coarse-to-fine and global-to-regional scheme ensures that this method is fast and accurate.

To further explore the localised eye centres, in Section 5, we introduce gaze gesture recognition, i.e. classification of the trajectory patterns of eye centre locations in consecutive frames. These gaze gestures provide contactless substitutes for mouse/keyboard input. Gaze gesture recognition therefore offers great assistance to the elderly and the disabled in accessing a variety of digital systems.

As our algorithms only require a standard webcam to capture image data, we implemented all the above mentioned algorithms and integrated them in a directed advertising system – which is one of the assistive technology tools our algorithms can bring to real-world applications. The directed advertising system combines the interpretation of demographic data (gender) and behavioural data (gaze) in order to deliver customised advertisements to its users and allow them to remotely browse advertising messages.

In summary, the main contribution of this paper is fourfold.

- (1) A novel gender recognition method utilising the Fisher Vector encoding method – a generic method that can encode almost all types of features but still maintains low complexity in implementation. It also proves to have high accuracy and robustness against head poses. To the best of our knowledge, this is the first time that Fisher Vectors have been employed for gender recognition.
- (2) A modular eye centre localisation method consisting of two modules and a Selective Oriented Gradient (SOG) filter. The eye centre localisation method has proved to be accurate, fast and, more importantly, robust to in-plane and out-of-plane head rotations. The SOG filter is specifically designed to detect and remove strong gradients from eyebrows, eye corners and self-shadows that many other methods suffer. The SOG filter also resolves general tasks regarding the detection of curved shapes.
- (3) Design of gaze gestures and a gaze gesture recognition algorithm. The algorithm captures the relative attention of the users and enables them to control HCI systems by issuing gaze gestures. This will largely enhance the interactivity of current HCI systems.
- (4) Development of an intelligent and assistive case study system. By combining demographic data and behavioural data, this system makes digital out-of-home advertising more adaptive and interactive so that it is of great value to both advertising agencies and consumers. This system also provides enabling technology for assisting diverse groups of people (including the elderly and the disabled), with accessing HCI systems with ease and convenience, in a wide range of applications.

Download English Version:

<https://daneshyari.com/en/article/6937553>

Download Persian Version:

<https://daneshyari.com/article/6937553>

[Daneshyari.com](https://daneshyari.com)