esture Recognition

Gesture recognition is an area of research and development in computer science and language technology concerned with the recognition and interpretation of human gestures. A subdiscipline of computer vision,[citation needed] it employs mathematical algorithms to interpret gestures.[1]

Gesture recognition offers a path for computers to begin to better understand and interpret human body language, previously not possible through text or unenhanced graphical user interfaces(GUI's).

Gestures can originate from any bodily motion or state, but commonly originate from the face or hand. One area of the field is emotion recognition derived from facial expressions and hand gestures. Users can make simple gestures to control or interact with devices without physically touching them.

Many approaches have been made using cameras and computer vision algorithms to interpret sign language, however, the identification and recognition of posture, gait, proxemics, and human behaviors is also the subject of gesture recognition techniques.[2]

Gesture recognition has application in such areas as:[when?]

Gesture recognition can be conducted with techniques from computer vision and image processing.[5]

The literature includes ongoing work in the computer vision field on capturing gestures or more general human pose and movements by cameras connected to a computer.[6][7][8][9]

The term "gesture recognition" has been used to refer more narrowly to non-text-input handwriting symbols, such as inking on a graphics tablet, multi-touch gestures, and mouse gesture recognition. This is computer interaction through the drawing of symbols with a pointing device cursor.[10][11][12] Pen computing expands digital gesture recognition beyond traditional input devices such as keyboards and mice, and reduces the hardware impact of a system.[how?]

In computer interfaces, two types of gestures are distinguished:[13] We consider online gestures, which can also be regarded as direct manipulations like scaling and rotating, and in contrast, offline gestures are usually processed after the interaction is finished; e. g. a circle is drawn to activate a context menu.

A touchless user interface (TUI) is an emerging type of technology wherein a device is controlled via body motion and gestures without touching a keyboard, mouse, or screen.[14]

There are several devices utilizing this type of interface such as smartphones, laptops, games, TVs, and music equipment.

One type of touchless interface uses the Bluetooth connectivity of a smartphone to activate a company's visitor management system. This eliminates having to touch an interface, for convenience or to avoid a potential source of contamination as during the COVID-19 pandemic.[15]

The ability to track a person's movements and determine what gestures they may be performing can be achieved through various tools. Kinetic user interfaces (KUIs) are an emerging type of user interfaces that allow users to interact with computing devices through the motion of objects and bodies.[citation needed] Examples of KUIs include tangible user interfaces and motion-aware games such as Wii and Microsoft's Kinect, and other interactive projects.[16]

Although there is a large amount of research done in image/video-based gesture recognition, there is some variation in the tools and environments used between implementations.

Depending on the type of input data, the approach for interpreting a gesture could be done in different ways. However, most of the techniques rely on key pointers represented in a 3D coordinate system. Based on the relative motion of these, the gesture can be detected with high accuracy, depending on the quality of the input and the algorithm's approach.[30]

In order to interpret movements of the body, one has to classify them according to common properties and the message the movements may express. For example, in sign language, each gesture represents a word or phrase.

Some literature differentiates 2 different approaches in gesture recognition: a 3D model-based and an appearance-based.[31] The foremost method makes use of 3D information on key elements of the body parts in order to obtain several important parameters, like palm position or joint angles. Approaches derived from it such as the volumetric models have proven to be very intensive in terms of computational power and require further technological developments in order to be implemented for real-time analysis. Alternately, appearance-based systems use images or videos for direct interpretation. Such models are easier to process, but usually lack the generality required for human-computer interaction.

The 3D model approach can use volumetric or skeletal models or even a combination of the two. Volumetric approaches have been heavily used in the computer animation industry and for computer vision purposes. The models are generally created from complicated 3D surfaces, like NURBS or polygon meshes.

The drawback of this method is that it is very computationally intensive, and systems for real-time analysis are still to be developed. For the moment, a more interesting approach would be to map simple primitive objects to the person's most important body parts (for example cylinders for the arms and neck, sphere for the head) and analyze the way these interact with each other. Furthermore, some abstract structures like super-quadrics and generalized cylinders maybe even more suitable for approximating the body parts.

Instead of using intensive processing of the 3D models and dealing with a lot of parameters, one can just use a simplified version of joint angle parameters along with segment lengths. This is known as a skeletal representation of the body, where a virtual skeleton of the person is computed and parts of the body are mapped to certain segments. The analysis here is done using the position and orientation of these segments and the relation between each one of them(for example the angle between the joints and the relative position or orientation)

Advantages of using skeletal models:

Appearance-based models no longer use a spatial representation of the body, instead deriving their parameters directly from the images or videos using a template database. Some are based on the deformable 2D templates of the human parts of the body, particularly the hands. Deformable templates are sets of points on the outline of an object, used as interpolation nodes for the object's outline approximation. One of the simplest interpolation functions is linear, which performs an average shape from point sets, point variability parameters, and external deformation. These template-based models are mostly used for hand-tracking, but could also be used for simple gesture classification.

The second approach in gesture detection using appearance-based models uses image sequences as gesture templates. Parameters for this method are either the images themselves, or certain features derived from these. Most of the time, only one (monoscopic) or two (stereoscopic) views are used.

Electromyography (EMG) concerns the study of electrical signals produced by muscles in the body. Through classification of data received from the arm muscles, it is possible to classify the action and thus input the gesture to external software.[1] Consumer EMG devices allow for non-invasive approaches such as an arm or leg band and connect via Bluetooth. Due to this, EMG has an advantage over visual methods since the user does not need to face a camera to give input, enabling more freedom of movement.

There are many challenges associated with the accuracy and usefulness of gesture recognition and software designed to implement it. For image-based gesture recognition, there are limitations on the equipment used and image noise. Images or video may not be under consistent lighting, or in the same location. Items in the background or distinct features of the users may make recognition more difficult.

The variety of implementations for image-based gesture recognition may also cause issues with the viability of the technology for general usage. For example, an algorithm calibrated for one camera may not work for a different camera. The amount of background noise also causes tracking and recognition difficulties, especially when occlusions (partial and full) occur. Furthermore, the distance from the camera, and the camera's resolution and quality, also cause variations in recognition accuracy.

In order to capture human gestures by visual sensors robust computer vision methods are also required, for example for hand tracking and hand posture recognition[32][33][34][35][36][37][38][39][40] or for capturing movements of the head, facial expressions or gaze direction.

One significant challenge to the adoption of gesture interfaces on consumer mobile devices such as smartphones and smartwatches stems from the social acceptability implications of gestural input. While gestures can facilitate fast and accurate input on many novel form-factor computers, their adoption and usefulness are often limited by social factors rather than technical ones. To this end, designers of gesture input methods may seek to balance both technical considerations and user willingness to perform gestures in different social contexts.[41] In addition, different device hardware and sensing mechanisms support different kinds of recognizable gestures.

Gesture interfaces on mobile and small form-factor devices are often supported by the presence of motion sensors such as inertial measurement units (IMUs). On these devices, gesture sensing relies on users performing movement-based gestures capable of being recognized by these motion sensors. This can potentially make capturing signals from subtle or low-motion gestures challenging, as they may become difficult to distinguish from natural movements or noise. Through a survey and study of gesture usability, researchers found that gestures that incorporate subtle movement, which appear similar to existing technology, look or feel similar to every action, and are enjoyable were more likely to be accepted by users, while gestures that look strange, are uncomfortable to perform, interfere with communication, or involve uncommon movement caused users more likely to reject their usage.[41] The social acceptability of mobile device gestures relies heavily on the naturalness of the gesture and social context.

Wearable computers typically differ from traditional mobile devices in that their usage and interaction location takes place on the user's body. In these contexts, gesture interfaces may become preferred over traditional input methods, as their small size renders touch-screens or keyboards less appealing. Nevertheless, they share many of the same social acceptability obstacles as mobile devices when it comes to gestural interaction. However, the possibility of wearable computers being hidden from sight or integrated into other everyday objects, such as clothing, allow gesture input to mimic common clothing interactions, such as adjusting a shirt collar or rubbing one's front pant pocket.[42][43] A major consideration for wearable computer interaction is the location for device placement and interaction. A study exploring third-party attitudes towards wearable device interaction conducted across the United States and South Korea found differences in the perception of wearable computing use of males and females, in part due to different areas of the body considered socially sensitive.[43] Another study investigating the social acceptability of on-body projected interfaces found similar results, with both studies labelling areas around the waist, groin, and upper body (for women) to be least acceptable while areas around the forearm and wrist to be most acceptable.[44]

Public Installations, such as interactive public displays, allow access to information and displays interactive media in public settings such as museums, galleries, and theaters.[45] While touch screens are a frequent form of input for public displays, gesture interfaces provide additional benefits such as improved hygiene, interaction from a distance, and improved discoverability, and may favor performative interaction.[42] An important consideration for gestural interaction with public displays is the high probability or expectation of a spectator audience.[45]

Arm fatigue was a side-effect of vertically oriented touch-screen or light-pen use. In periods of prolonged use, users' arms began to feel fatigued and/or discomfort. This effect contributed to the decline of touch-screen input despite its initial popularity in the 1980s.[46][47]

In order to measure arm fatigue side effect, researchers developed a technique called Consumed Endurance.[48][49]