Computer Vision and Image Understanding 141 (2015) 152–165

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Recent methods and databases in vision-based hand gesture recognition: A review

Pramod Kumar Pisharady*, Martin Saerbeck

*Institute of High Performance Computing, A*STAR, #16-16 Connexis, 1 Fusionopolis Way, Singapore 138632, Singapore*

ARTICLE INFO

Article history:
Received 16 October 2014
Accepted 11 August 2015

Keywords:
Gesture recognition
Posture recognition
Hand pose estimation
Sign language recognition
Hand gesture dataset
Gesture database
Survey

ABSTRACT

Successful efforts in hand gesture recognition research within the last two decades paved the path for natural human–computer interaction systems. Unresolved challenges such as reliable identification of gesturing phase, sensitivity to size, shape, and speed variations, and issues due to occlusion keep hand gesture recognition research still very active. We provide a review of vision-based hand gesture recognition algorithms reported in the last 16 years. The methods using RGB and RGB-D cameras are reviewed with quantitative and qualitative comparisons of algorithms. Quantitative comparison of algorithms is done using a set of 13 measures chosen from different attributes of the algorithm and the experimental methodology adopted in algorithm evaluation. We point out the need for considering these measures together with the recognition accuracy of the algorithm to predict its success in real-world applications. The paper also reviews 26 publicly available hand gesture databases and provides the web-links for their download.

1. Introduction

Nonverbal communication, which includes communication through hand gestures, body postures, and facial expressions makes up about two-thirds of all communication among human [1]. Hand gestures are one of the most common category of body language used for communication and interaction. Whilst the rest of the body indicates a more general emotional state, hand gestures can have specific linguistic content in it [2]. Due to the speed and expressiveness in interaction, hand gestures are widely used in sign languages and human–computer interaction systems.

One ongoing goal in human–machine interface design is to enable effective and engaging interaction. For example, vision-based hand gesture recognition (HGR) systems can enable contactless interaction in sterile environments such as hospital surgery rooms, or simply provide engaging controls for entertainment and gaming applications. However HGR is not as robust as standard keyboard and mouse based interaction. Issues such as sensitivity to size and speed variations, poor performance against complex backgrounds and varying lighting conditions, and the reliable detection of gesturing phase have limited the use of hand gestures as a reliable modality in interface design.

1.1. Taxonomy of gestures

There are multiple ways to categorize hand gestures, (1) based on observable features and (2) based on the interpretation. In the first category gestures are classified based on temporal relationships, into two types; *static* and *dynamic* gestures (Fig. 1). Static hand gestures (*aka* hand postures/hand poses) are those in which the hand position does not change during the gesturing period. Static gestures mainly rely on the shape and flexure angles of the fingers. In dynamic hand gestures, the hand position changes continuously with respect to time. Dynamic gestures generally have three motion phases: preparation, stroke, and retraction [3]. The message in a dynamic gesture is mainly contained in the temporal sequence in the stroke phase. Dynamic gestures rely on the hand trajectories and orientations, in addition to the shape and fingers' flex angles.

In the second category, gestures are classified based on the interpreted meaning. For example emblems, illustrators, regulators, affect displays, and adaptors [4,5] are the typical classes to describe gestures. Emblems (also labeled as autonomous gestures) are gestures that can be substituted for spoken words (for example, showing *thumbs-up* instead of saying *all right*). Illustrators are gestures used to illustrate spoken words (for example, giving directions by *pointing*). Regulators support the interaction and communication between speaker and listener (for example, raising hand to manage turn-taking). Affect displays are facial expressions, which when combined with postures reflect the intensity of an emotion (for example, staring at an object and moving the body back reflect the emotion *fear*). Adaptors are gestures used at some point in time for

* Corresponding author.
E-mail addresses: pramodkumar@ieee.org, pramodkp@mit.edu (P.K. Pisharady), saerbeckm@ihpc.a-star.edu.sg (M. Saerbeck).

http://dx.doi.org/10.1016/j.cviu.2015.08.004
1077-3142/

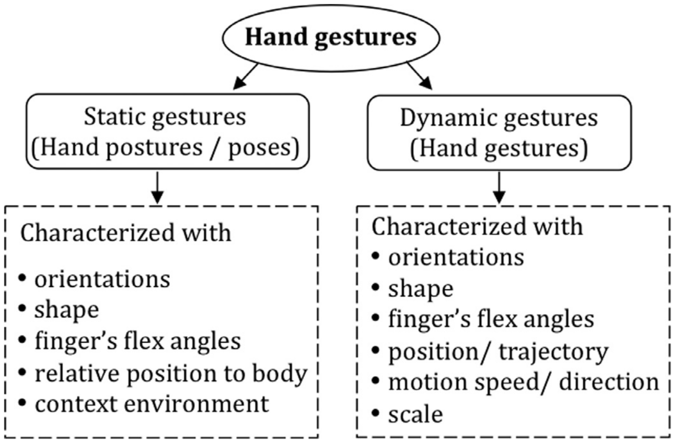

Fig. 1. Classification of hand gestures based on temporal nature. Static gestures are time independent whereas dynamic gestures are time dependent.

personal convenience, but have turned into a habit (for example, adjusting glasses in a tensed situation).

1.2. Hand gesture recognition

Fig. 2 shows the block diagram of a typical contactless gesture recognition system. The sensor is a camera in vision-based gesture recognition systems. Berman et al. [6] reviewed different sensors used in gesture recognition systems and provided a comprehensive analysis of integration of sensors into gesture recognition systems and their impact on the system performance. Based on feature extraction, vision-based gesture recognition systems are broadly divided into two categories, appearance-based methods and three dimensional (3D) hand model-based methods. Appearance-based methods utilize features of training image to model the visual appearance, and compare these parameters with the features of test image. Three-dimensional model-based methods rely on a 3D kinematic model, by estimating the angular and linear parameters of the model.

1.3. Survey and evaluation of hand gesture recognition techniques

Our study builds on top of earlier attempts to survey the field of HGR. Mitra et al. [7] provided a survey of different gesture recognition methods, covering hand and arm gestures, head and face gestures, and body gestures. The HGR methods investigated in the survey was limited to Hidden Markov Models (HMMs), particle filtering and condensation algorithms, and Artificial Neural Networks (ANNs). Hand modeling and 3D motion based pose estimation methods are reviewed in [8] (ignoring the gesture classification schemes). An analysis of sign languages, grammatical processes in sign gestures, and issues relevant to the automatic recognition of sign languages are discussed in [9]. The latest of the above papers [8] covered developments

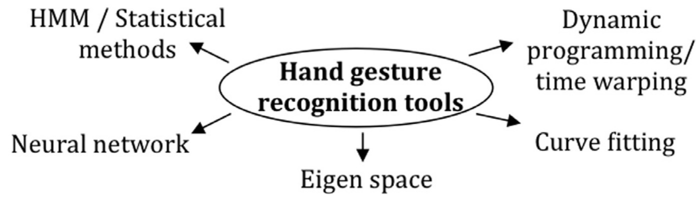

Fig. 3. Taxonomy of hand gesture recognition techniques reviewed.

till the year 2005. The review concluded that the methods studied are experimental and their use is limited to laboratory environments.

This paper reviews recent works in HGR with a focus on the developments in the last 16 years. Algorithms utilizing conventional RGB cameras (Section 2) as well as the new generation RGB-D cameras (Section 3) are surveyed, making the review unique. The HGR methods are classified and analyzed according to the technique used for gesture classification. We perform a quantitative comparison of HGR algorithms based on different attributes of the algorithm and the experimental methodology followed in algorithm testing. A review of available hand gesture databases (Section 4) and a discussion on hand gesture recognition research (Section 5) are also provided. We hope this survey is timely, given the growing research efforts and expanding market for gestural interactive systems.

2. Conventional hand gesture recognition: RGB sensor based methods

2.1. Recognition of dynamic hand gestures

The techniques used for dynamic HGR can be classified as (a) HMM [10–23] and other statistical methods [24–31], (b) ANN [32–34] and other learning based methods [35,36], (c) Eigenspace based methods [37,38], (d) Curve fitting [39], and (e) Dynamic programming [40]/Dynamic time warping [41,42] (Fig. 3).

2.1.1. HMM and other statistical methods

HMM is the most widely used HGR technique. HMM is a statistical model in which the system being modeled is assumed to be a Markov process with unknown parameters. HMM represents the statistical behavior of an observable symbol sequence using a network of hidden states with transition and emission probabilities. The HMM can be used for pattern recognition once the hidden parameters are identified using the observable data.

HMM based dynamic hand gesture recognition methods mainly utilize temporal and spatial features of input images. Chen et al. [14] utilized Fourier descriptor and optical flow based motion analysis to characterize spatial and temporal features respectively. The algorithm extracts hand shape from complex backgrounds by tracking the hand in realtime. HMM based recognizers identify the best likelihood gesture model for a given pattern. The variations in gesture from a reference pattern reduce the likelihood of the gesture with the model.

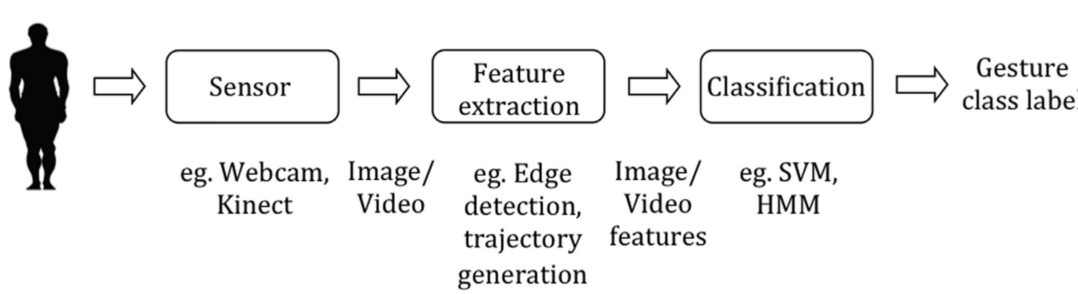

Fig. 2. Gesture recognition pipeline.