Engineering Applications of Artificial Intelligence 35 (2014) 54–70

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Real time hand detection in a complex background

Ekaterini Stergiopoulou, Kyriakos Sgouropoulos, Nikos Nikolaou, Nikos Papamarkos *, Nikos Mitianoudis

Image Processing and Multimedia Laboratory, Department of Electrical & Computer Engineering, Democritus University of Thrace, 67100 Xanthi, Greece

ARTICLE INFO

Article history:
Received 17 October 2013
Received in revised form
12 February 2014
Accepted 13 June 2014
Available online 8 July 2014

Keywords:
Region based hand detection
Complex background
Hybrid motion detection
Skin color detection
Color reduction

ABSTRACT

Hand gesture recognition has gained the interest of many researchers in recent years, as it has become one of the most popular Human Computer Interfaces. The first step in most vision-based gesture recognition systems is the hand region detection and segmentation. This segmentation can be a particularly challenging task when it comes to complex backgrounds and varying illumination. In such environments, most hand detection techniques fail to obtain the exact region of the hand shape, especially in cases of dynamic gestures. Meeting these requirements becomes even more difficult, due to real-time operation demand. To overcome these problems, in this paper, we propose a new method for real-time hand detection in a complex background. We employ a combination of existing techniques, based on motion detection and introduce a novel skin color classifier to improve segmentation accuracy. Motion detection is based on image differencing and background subtraction. Skin color detection is accomplished via a color classification technique that employs online color training, so that the system can dynamically adapt to the variety of lighting conditions and the user's skin color as well as possible. Morphological features of the detected hand in previous frames are employed to estimate the probability of a pixel belonging to the hand section in the current frame. Finally, the derived motion, color and morphological information are combined to detect the hand region. Experimental results show significant improvement in hand region detection, compared to existing methods with an average accuracy of 98.75%.

1. Introduction

The increasing spread of intelligent computing in everyday life has introduced a growing need for more intuitive and efficient ways of interaction between human and computers. Hand gestures are an appealing alternative to traditional currently-used devices (keyboard, mouse), since they form an extensive part of natural human communication (Ebert et al., 2012). Vision-based gesture recognition provides the potential for creating a new, easier and more powerful human – computer interface, because this task does not require any special hardware that might hinder user's comfort. It is also a non-intrusive information processing tool with many capabilities and a low-cost method, since only a web camera is required. Vision-based hand gesture recognition has a range of applications, such as sign language interpretation and learning, teleconferencing, distance learning, robotics, games, selection and object manipulation in virtual environments (Wachs et al., 2011).

Hand gesture recognition systems commonly consist of three main stages: (i) hand detection, (ii) hand feature extraction, and (iii) gesture recognition. Two major difficulties are usually encountered in these systems:

- Uncontrolled environments: An ideal hand gesture recognition system should operate regardless of the background complexity or the variety of lighting conditions. Nevertheless, the task of locating a rigid object in a complex background remains challenging in computer vision (Erol et al., 2007).
- Processing speed: The systems should be able to perform real-time gesture recognition. If the system performance is slow, it will be unacceptable for commercial applications. Simple and computationally efficient features are of great interest to machine vision (Wachs et al., 2011).

A common technique to cope with these difficulties is to apply restrictions on the user or the environment (Chua et al., 2002; Flasiński and Myśliński 2010; Ge et al., 2008; Lee 2008; Shimada et al., 2001; Ueda et al., 2003; Vatavu et al., 2009; Wilkowski 2009; Yang et al., 2009; Zhao and Chen 2009). Commonly encountered assumptions are that the background is plain, the

* Corresponding author: Tel.: +30 25410 79585; fax: +30 25410 79569.
E-mail address: papamark@ee.duth.gr (N. Papamarkos).
URL: http://www.papamark.gr/ (N. Papamarkos).

http://dx.doi.org/10.1016/j.engappai.2014.06.006
0952-1976/

hand is the only skin-colored object in the observed scene and that the lighting conditions are specific. The same assumptions were used in our earlier work (Stergiopoulou and Papamarkos 2009) on a hand-gesture recognition system.

In order to address the problem of vision-based hand detection in a complex background, most approaches in the literature employ visual features such as color, motion information, shape or a combination of these (Zabulis et al., 2009). Skin color can be estimated using off-line training data or face-pixel colors. In some cases, the skin color model can be adapted using the detected hand pixels from previous frames. Motion detection is achieved mainly by image differencing or/and background subtraction algorithms. The most representative methods for hand detection can be divided in two main categories: (i) methods that estimate a region containing the hand, and (ii) methods that extract the exact shape of the hand.

For hand region detection, Dadgostar et al. (2009) employ a thresholding skin detector in the hue color space, whose thresholds are constantly recalculated using the moving skin pixels of the scene. Moving skin pixels are detected by image differencing between consecutive frames. This method performs well as long as non-skin objects appear in the scene, whose hue color component falls into the skin detector range. Wilson and Salgian (2008) implement a gesture recognition method that uses a background subtraction technique. However, they assume that the background would be static with no illumination variation. Face and hand region detection is achieved by applying a Bayes classifier and Gaussian mixture models for the skin and non-skin classes. Alon et al. (2005) detect the face and then use the mean and the covariance of the face skin pixels colors in the normalized RGB color space to compute the skin likelihood image. If there is significant motion between the previous and the current frame, a motion mask, produced by image differencing, is applied to the skin likelihood image so as to estimate the hand likelihood image. They use sub-windows to extract the hand region based on the sum of their pixels likelihood. Their method implies that user's face is present in the scene, ideally illuminated, in order to construct the skin likelihood model. Guo et al. (2012) use an object detector based on weak classifiers, hard-thresholding skin color segmentation in HSV color space and background cancelation for hand region detection under complex background. Their method reduces the training time of their detector by using a new set of pixel-based hierarchical-features. The proposed window-based features exploit the fact that the hand is centered in each of the training images. Background cancelation is based on a pixel-wise background model trained over a period of time. In general, background cancellation improves hand detection. However, the hard-thresholding skin color classifier in the HSV color space cannot deal effectively with false positive detections.

For hand shape extraction, Alexander et al. (2009) use a two-frame difference to identify areas containing motion. Also, they perform hand detection by using a corner detection algorithm and geometric features of the hand. Unfortunately, the authors do not provide experimental results to evaluate their method and the presented techniques have not been examined in detail. Zhao et al. 2008 implement real-time gesture segmentation based on dual-complexion and an adaptive complex background model. Initially, they build the complexion model by means of a Gaussian distribution in the YCbCr color space and then apply it on the input image. The results are refined via a thresholding skin detection technique in the normalized RGB color space. The authors propose a background adapting modeling technique also based on the Gaussian distribution, which is able to adapt environment changes, but fails when the hand overlaps with skin color background. Chen et al. 2003 detect the moving region by image differencing. The result is refined by comparing the non-moving objects with the sample skin color. Then, edge detection is applied in order to separate the arm from the hand. The segmented hand is the output of the bitwise logical "AND" operation on the motion, skin and edge detection results. They also use a background subtraction technique, with a continuously updated background. Their method could lead to false positives, when a large object, like a sleeve, moves with the hand, since they do not use any pre-trained color model, but only color samples from moving objects. The HSV color space is used by Dardas and Georganas (2011) for thresholding classification of skin and non-skin regions. The contours of the skin detected regions are then compared to the contours of static gestures templates to decide whether the detected region contains a hand or it is a false positive. However, the authors do not provide any details on the actual contour comparison algorithm, and on the impact of the number of different gesture templates on the detection rate. Donoser and Bischof (2008) combined a skin color likelihood algorithm with an interest region detector for real-time hand tracking. They analyze color cues to calculate a skin color probability value for every pixel in the frame. A detector that estimates the high probability-connected regions, which display low probability values along their boundaries, is then applied to extract the hand region. The proposed technique can be applied only when the hand is the only object that is similar to the color model. Mao et al. (2009) combine the object detector, proposed by Viola and Jones (2001), with a skin color filtering technique to detect and track the hand in complex background. Skin detection is applied to remove background pixels, followed by a real-time object detector. The hand detector introduced by this technique improves the standard object detector of Viola and Jones against complex background, but still fails with skin color background. Okkonen et al. (2007) combine background subtraction with histogram-based color segmentation for a robust skin area segmentation algorithm. The main disadvantage of this method is that it cannot adapt to changes in the background since the background image is composed as an average of the first N images in the video sequence.

In this paper we propose a new method for real-time hand detection in a complex background. The main motivation behind the proposed technique is to address the problem of uncontrolled environments without using restrictions along with low computational cost and inexpensive hardware. These difficulties can be amended by means of a robust and more efficient hand detection technique. The proposed real-time hand detection method takes advantage of motion, skin color and morphology information, in order to increase effectiveness and robustness. The aim of our paper is the implementation of an effective and real-time hand detection system which operates in a complex background and under various illumination conditions. In addition, this system could be exploited further for dynamic gesture recognition. The second objective of the proposed technique is the precise extraction of the hand shape, i.e. the palm and the raised fingers should be well detected by the system, aiming further at static gesture recognition.

The novelties introduced in this paper are summarized as follows: Firstly, we introduce a modification to the motion detection algorithm of Collins et al. (2000). The proposed modification addresses the problem of misdetection when the moving object has the same color as the background, a common situation in hand detection applications. We also propose the combination of online and offline training of the Skin Color Map (Bayes) classifier, which has been used by other researchers only as a pre-trained classifier. Additionally, an algorithm which defines morphology weights of hand pixels is proposed. Finally, our technique employs a color reduction algorithm to define arbitrary shaped areas of similar color in which the derived motion, color and morphological information is combined. The proposed region-based approach differs significantly from other methods mentioned earlier, as we