Implementing a Visual Servoing System for Robot Controlling

Maryam Vafadar , Alireza Behrad and Saeed Akbari

***Abstract*—**Nowadays, with the emerging of the new applications like robot control in image processing, artificial vision for visual servoing is a rapidly growing discipline and Human-machine interaction plays a significant role for controlling the robot. This paper presents a new algorithm based on spatio-temporal volumes for visual servoing aims to control robots. In this algorithm, after applying necessary pre-processing on video frames, a spatio-temporal volume is constructed for each gesture and feature vector is extracted. These volumes are then analyzed for matching in two consecutive stages. For hand gesture recognition and classification we tested different classifiers including k-Nearest neighbor, learning vector quantization and back propagation neural networks. We tested the proposed algorithm with the collected data set and results showed the correct gesture recognition rate of 99.58 percent. We also tested the algorithm with noisy images and algorithm showed the correct recognition rate of 97.92 percent in noisy images.

***Keywords*—**Back propagation neural network, Feature vector, Hand gesture recognition, k-Nearest Neighbor, Learning vector quantization neural network, Robot control, Spatio-temporal volume, Visual servoing

I. Introduction

FOR years, scientists have been trying to teach machines how to see like humans, Nowadays Artificial vision and visual servoing is developed and Recent research has seemed to show advanced applications in different areas[1-3]. One of these emerging applications is robot control. Artificial vision for visual servoing is a rapidly growing discipline and human-machine interaction plays a significant role for interaction of human with a robot. Human hand gesture recognition is one of the mostly used methods for human-machine interaction and can be used as a perfect method for controlling a robot. Different vision-based hand gesture recognition methods have been proposed including applying Fourier transform [4] or Principal Component Analysis (PCA) [5-7] on images. Edge orientation histogram is also used for hand gesture recognition [8]. Using temporal templates is another method for hand gesture recognition [9-11]. The basic idea of temporal template is the projection of the temporal pattern of motion into a single image-based representation and the extraction of the appropriate features from this image.

Maryam Vafadar Department of Electronic, Sama technical and vocational training college, Islamic Azad University, Islamshahr Branch, Islamshahr, Iran, E-mail: vafadar_ma@yahoo.com

Alireza Behrad Department of Electronic, Shahed University, Tehran, Iran, E-mail: behrad@shahed.ac.ir

Saeed Akbari Department of Mathematics, Islamic Azad University, Tehran, Iran, E-mail: sa_akbari1@yahoo.com

In the most of these methods, video frames are primarily processed in the groups of two consecutive frames. For example in temporal template methods, by differentiating one frame from the other, the dynamics occurring in an image sequence are extracted. Edge orientation histogram is another pair-wise approach that recognizes gestures by extracting edge points as features in each frame and matching them between frames. Although the two-frame-based approaches have been very successful in some applications, they face considerable difficulties; if they used for example to reason about non-constant velocity motion [12]. Also, the presence of noise and occlusion affects feature extraction and correspondence stages and degrades the performance of the algorithms. Spatio-temporal volumes have been actively in research as a means to alleviate the shortcoming of the traditional pair-wise approaches. In fact, video sequence can be defined as a spatial intensity pattern that changes with time. Spatio-temporal volume unifies the analysis of spatial and temporal information by constructing a volume of data in which consecutive images are stacked to form a third, temporal dimension. Then processing of an image sequence is done as a 3D volume and required features are extracted from this volume.

The major advantages of this representation are as follows [12]:

1. By conjointly providing spatial and temporal continuity, the complexity of feature correspondence is significantly reduced.
2. Noise in each frame has low effect in the volume and rarely can degrade the performance of algorithm.
3. By analyzing feature structures in this volume, we may reason about non-constant velocities, efficiently.
4. Occlusion events are made much easier to detect, as they are represented explicitly in this volume as truncated paths [13-14].

Several spatio-temporal and spatio-temporal frequency representations have been proposed such as extracting geometric features from spatio-temporal volume [15], applying the derivative of Gaussian transform [16] and Fourier transform [17] on spatio-temporal volume.

This paper presents a new spatio-temporal algorithm for hand gesture recognition aims to control robots. In this algorithm, after applying necessary pre-processing on video frames, spatio-temporal volume is constructed for each gesture.

These volumes are then analyzed and feature vectors are extracted. In comparison with other method, we have used contour of the hand for volume construction. This reduces the computation overhead of the algorithm and make the hand gesture recognition algorithm less or insensitive to illumination change, hand size and hand color.