Neural and Statistical Methods for Adaptive Color Segmentation - A Comparison

Enno Littmann
Neural Information Processing
Computer Science Faculty
University of Ulm, D-89069 Ulm, FRG
enno@neuro.informatik.uni-ulm.de

Helge Ritter
Neuroinformatics, Computer Science Dept.
Bielefeld University, POB 100 131
D-33501 Bielefeld, FRG
helge@techfak.uni-bielefeld.de

Abstract

For a long time pixel based segmentation methods were restricted to grayscale images due to the enormous computational costs when dealing with color. With the availability of more powerful computers it is nowadays possible to perform pixel based operations on real camera images even in the full color space. New adaptive classification tools like neural networks make it possible to develop special-purpose object detectors that can segment arbitrary objects in real images with a complex distribution in the feature space after training with one or several previously labelled image(s). The proposed adaptive segmentation method uses local color information to estimate the membership probability in the object resp. background class. The method is applied to the recognition and localization of human hands in color camera images of complex laboratory scenes. The paper focuses on the influence of the chosen color representation and a detailed comparison of the neural approach to standard statistical methods, a threshold filter, and a classifier based on normal distributions.

1 Introduction

In the last years research and development of neural networks have received growing attention. Most investigations concern function approximation and classification tasks. Major objectives are the analysis of optimal features and learning from examples. These tasks as well as the objectives are well-known from pattern recognition [8]. From statistics we already know a number of "classical" methods that also deal with the problem of feature extraction and learning, and there are also knowledge-based approaches. These three fields, however, derive their methods from different paradigms [14]. Therefore, several methods from statistics have been re-invented in the context of neural networks [15, 14]. This close similarity requires a detailed architectural and performance comparison especially between neural and statistical methods.

In a collaboration at Bielefeld University we investigated a hybrid system to combine the advantages of knowledge-based and neural methods [3, 4, 2]. The system is based on local linear maps as neural networks (sec. 2.2) and the semantic network ERNEST [13, 5]. We applied this system to a challenging and important problem in pattern recognition, the recognition of human hands in real camera images. The solution of this task forms the basis for a number of subsequent processing steps during which further geometrical features, such as hand orientation and information on finger posture, are extracted. The development of a system with these capabilities is directed towards the realization of a more flexible and powerful man-machine-interface for the control of multi-fingered anthropomorphic manipulators or virtual reality environments by human gestures.

G. Sagerer et al. (eds.), Mustererkennung 1995
© Springer-Verlag Berlin Heidelberg 1995
This paper focuses on the analysis of one important step, namely the segmentation of the hand from the raw pixel images. The success of this step is vital, as the performance of the following image processing depends crucially on its result. We realize the segmentation as a local, pixel based classification of image pixels in "hand" and "background" by a neural network that was previously adapted with preclassified data. The proposed adaptive segmentation method uses local color information to estimate the membership probability in the object resp. background class. The method is applied to the recognition and localization of human hands in color camera images of complex laboratory scenes. The neural approach is compared to standard statistical methods, namely the box filter and a classifier based on normal distributions. We investigate the influence of the chosen feature space and of the training sample as well as the use of resolution pyramids. Further detailed investigations with additional data, including the performance evaluation for hands of different persons, can be found in [6, 7].

2 Classification methods and features

2.1 Color representations

So far, the \textit{RGB} system is the dominating representation of color due to its importance for television and camera systems. Physiologically, this representation is similarly realized in the human retina that consists of three color-sensitive photoreceptor types with a maximum spectral sensitivity corresponding to red, green, and blue. Besides the \textit{RGB} system there is a variety of other representations that can be constructed by linear or nonlinear transformations from the \textit{RGB} values. A survey can be found in [18] \footnote{Official definitions of color spaces and their transformations are given in [1].}.

One major disadvantage of the \textit{RGB} system is the dependency of all three parameters from the light intensity. This disadvantage is avoided in other representation schemes where color and intensity are represented independently. One such system is the \textit{Yuv} color space described in fig. 1. The color plane forms an equilateral triangle that can be constructed in the \textit{RGB} cube by connecting its \textit{R}, \textit{G}, and \textit{B} corner (fig. 2 (left)). The intensity \textit{Y} is perpendicular to the image plane. The values \textit{u} and \textit{v} are the cartesian coordinates of the color value. The formulation of this system in polar coordinates is called \textit{HSI} system. There we distinguish \textit{hue}, specified by the angle, \textit{saturation} as radial component, and \textit{intensity} again vertical to the \textit{uv} plane.

We prefer the \textit{Yuv} system because of its continuous representation of the color components. A similar effect of intensity independence can be achieved by simply calculating \textit{color ratios} \( Q_{RG} = R/(R + G) \) and \( Q_{RB} = R/(R + B) \).