Applied Soft Computing 46 (2016) 753–766

Contents lists available at ScienceDirect

Applied Soft Computing

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

Stereo-camera-based object detection using fuzzy color histograms and a fuzzy classifier with depth and shape estimations

Chia-Feng Juang, Guo-Cyuan Chen, Chung-Wei Liang, Demei Lee*

Department of Electrical Engineering, National Chung Hsing University, Taichung 402, Taiwan, ROC

ARTICLE INFO

Article history:
Received 11 April 2015
Received in revised form 6 October 2015
Accepted 12 October 2015
Available online 19 October 2015

Keywords:
Object detection
Fuzzy classifier
Disparity map
Shape detection
Object localization

ABSTRACT

This paper proposes a new method of detecting an object containing multiple colors with non-homogeneous distributions in complex backgrounds and subsequently estimating the depth and shape of the object using a stereo camera. To extract features for object detection, this paper proposes fuzzy color histograms (FCHs) based on the self-splitting clustering (SSC) of the hue-saturation (HS) color space. For each scanning window in a pyramid of scaled images, the FCH is obtained by accumulating the fuzzy degrees of all of the pixels belonging to each cluster. The FCH is fed to a fuzzy classifier to detect an object in the left image captured by the stereo camera. To find the matched object region in the right image, the left and right images are first segmented using the SSC-partitioned HS space. The depth of the object is then found by performing stereo matching on the segmented images. To find the shape of the object, a disparity map is built using the estimated object depth to automatically determine the stereo matching window size and disparity search range. Finally, the shape of the object is segmented from the disparity map. The experimental results of the detection of different objects with depth and shape estimations are used to verify the performance of the proposed method. Comparisons with different detection and disparity map construction methods are performed to demonstrate the advantage of the proposed method.

1. Introduction

The detection and localization of a specific object in an image is an important task in various computer-vision-based applications such as grasping an object using a robot arm [1,2], robot navigation [3,4], and image retrieval [5–7]. The detection of a specific object usually consists of the process of feature extraction followed by classification. Various feature extraction approaches have been proposed. One approach is to extract the gradient information of the appearance. Examples are the use of the Haar-like wavelet transformation method to extract the edge information in different orientations [8–11] and the scale-invariant feature transform (SIFT), which locates keypoints based on local image gradients [4,11,12]. These methods may be unsuitable for the detection of certain types of objects in complex environments [13]. Another approach is using histograms of pixels in a color space. Histograms obtained from grids with a uniform partition [6,14,15] or non-uniform partition [16] of a color space have been proposed, where the latter has been shown to provide better performance compared to the former [16]. In contrast to grid-type partitioning, flexible partitioning of the color space using a self-splitting clustering (SSC) algorithm has been proposed [13], where the entropy of color components based on histograms of pixels in different bins (clusters) is extracted as a detection feature. In these studies, histograms are obtained by counting the total number of pixels located in each grid/cluster.

In contrast to the traditional histogram computation approach introduced above, this paper proposes fuzzy color histograms (FCHs), whereby histograms are obtained based on the accumulated fuzzy degrees of all the pixels belonging to each cluster of an SSC-partitioned color space. The FCH is fed to a fuzzy classifier (FC) to detect an object. In terms of the classification method, the template matching method has been widely used in several studies for object detection [4,11,15]. In this method, a small set of templates would cause poor generation performance. A large set of templates help improve the generalization performance at the cost of increased memory requirements and computational load. FCs have been shown to provide a good classification performance with a small model size [16–20]. This paper uses an FC learned through a support vector machine (SVM) because the SVM-based FCs have been shown to outperform neural fuzzy classifiers and

* Corresponding author. Department of Mechanical Engineering, Chang Gung University, Taiwan, ROC. Fax: +886 3 2118050.
E-mail address: dmlee@mail.cgu.edu.tw (D. Lee).

http://dx.doi.org/10.1016/j.asoc.2015.10.025
1568-4946/

SVMs in different classification problems [13,16,20–22], such as object classification [13,16] and skin-color segmentation problems [20,21].

This paper uses a stereo camera to detect an object containing multiple colors with non-homogeneous distributions in complex backgrounds and subsequently estimate its depth and shape. For this type of object, it would be infeasible to find the shape of the object using color segmentation, a method generally applied to objects containing a single color in a plain background. The camera simultaneously captures left and right images. An object is detected from the left image using the FCH-based FC. The depth and shape of the detected object are then estimated based on the disparity map constructed from the stereo matching of the left and right images. Based on the format of the signals used in the stereo matching process, stereo matching algorithms can be divided into direct methods and indirect methods [23–32]. The direct method uses the original image intensity, such as the gray level [23,24,26] and color value [25,32], for stereo matching. The indirect method transforms the original image intensity into another feature [24,27–31] such as gradients of pixels [24,27]. In real scenarios, due to radiometric variations in a stereo image pair, two corresponding pixels will rarely share the same pixel values. Under such conditions, the direct matching method may suffer from severe performance degradation when the variations in the colors or gray levels are significant. To reduce the influence of the radiometric variations, this paper uses the indirect method. The SSC-partitioned color space is used to segment the original color image, where a cluster in the color space corresponds to a segmented color region. That is, the number of segmented color regions is equal to the number of clusters. The segmentation operation shares the same partitioned color space as the detection operation, which reduces the segmentation time. Stereo matching is performed on the segmented image to improve the performance of estimating the depth and shape of the detected object.

The contribution of this paper is twofold. First, this paper proposes a new histogram-based feature FCH to detect an object in an image. The improved performance of the FCH-based FC detection method is verified through comparisons with various histogram-based classification methods. Second, most previous studies on object detection consider only the detection of an object in an image and use a rectangular box to locate the object [8–16]. In this paper, based on the 2D location of the detected object in an image, a new method is proposed to estimate the depth and shape of the object using a new disparity map construction method. The use of SSC-based image segmentation and object-depth-based adaptive window size matching techniques is proposed to improve the estimation performance. The improved performance of the proposed image representation method is verified through comparisons with various direct and indirect methods.

This paper is organized as follows. Section 2 describes the FC. Section 3 introduces the SSC-based FCH feature extraction and object detection processes. Section 4 introduces the disparity map construction process and the estimation of the object depth and shape based on the map. Section 5 presents the experimental results of the detection and depth and shape estimation of three different objects. Performance comparisons of the proposed method with various histogram-based object detection and disparity map construction methods are conducted in this section. Finally, Section 6 presents the conclusions.

2. Fuzzy classifier

The FC is composed of fuzzy if-then rules in the following form:

$$\text{Rule i}: \quad \text{if } x_1 \text{ is } A_{i1} \text{ and}, \ldots, \text{ and } x_n \text{ is } A_{in} \text{ then } y' \text{ is } a_i \tag{1}$$

where a_i is a crisp value and A_{ij} is a fuzzy set with a Gaussian membership function described as follows:

$$M_{ij}(x_j) = \exp\left\{-\frac{(x_j - m_{ij})^2}{\sigma_i{}^2}\right\} \tag{2}$$

where m_{ij} and σ_i denote the center and width of the fuzzy set, respectively. Using the algebraic product for the t-norm "AND" operation and the weighted sum for the defuzzification operation, the FC output is

$$y' = \sum_{i=1}^{r} a_i \cdot \prod_{j=1}^{n} M_{ij}(x_j) + b \tag{3}$$

where b is a classification bias determined after parameter learning through SVM.

The antecedent and consequent parameters of the FC are learned through SSC and a linear SVM, respectively, as in [13]. Given a training data set, the SSC algorithm starts with one cluster and generates new clusters through a splitting operation. Among r existing clusters, a cluster is selected for splitting using a given selection criterion. In designing the FC, the applied criterion is the maximum variance of the samples in a cluster so that all clusters show similar variances after clustering. The selected cluster is split into two new clusters whose initial centers are assigned as the two training samples that have the minimum Euclidean distances to the original cluster center $\vec{m}_l$. After splitting, new centers of these $r+1$ clusters are recomputed by a k-means clustering algorithm. The splitting and k-means clustering operations repeat until a predefined cluster number, $C_{\max}$, is met. After the SSC clustering, the number of clusters is equal to the number of rules in the FC. The center and standard deviation of each cluster are assigned as the fuzzy set center and width of the fuzzy rule. The SSC algorithm avoids the random initialization of cluster centers in the traditional k-means clustering. Cluster splitting using the variance criterion enables similar color patterns to belong to the same cluster and the input space covered by a cluster (rule) to be moderate so that the generated fuzzy sets can properly cover the domain of each input variable. One popular approach of generating fuzzy rules in SVM-based FCs is using Gaussian-kernel-based support vectors (SVs) [33,34], where an SV corresponds to a fuzzy rule. For multi-color object classification problems, this approach suffers from the large size of a classification model because the number of SVs in an SVM is usually very large [13,16]. The above descriptions explain the suitability of using the SSC algorithm in the FC-based multi-color object detection problem.

After the determination of the antecedent parameters, the consequent parameters a_i of the FC are determined using a linear SVM, the details of which can be found in [13]. Neural fuzzy systems tune the consequent parameters based on training-error minimization, which considers only empirical risk error minimization and does not consider the overtraining problem. Formulation of SVMs is based on the principle of structural risk minimization [35]. Tuning of the consequent parameters using SVM considers not only the minimization of training error but also the separation margin between two classes. As stated in [35,Ch. 4], the bound of generalization error depends not only on training error but on classification margin as well. That is, the use of a linear SVM reduces the bound on test error, and therefore, reduces the chance of overtraining.

3. Object detection using fuzzy color histograms

This paper considers object detection and depth and shape estimation using a stereo red-green-blue (RGB) camera. The two lenses in the stereo camera are placed with their optical directions in parallel. The stereo camera consists of two cameras combined to