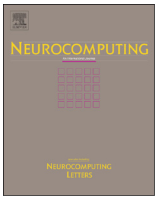
Neurocomputing 181 (2016) 116–121

Contents lists available at ScienceDirect

Neurocomputing

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



Human articulated body recognition method in high-resolution monitoring images

Yi Li [a,b], Xun Liu [b], Sanyuan Zhang [b], Xiuzi Ye [a]

[a] *College of Mathematics and Information Science, Wenzhou University, Wenzhou 325035, China*
[b] *College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*

ARTICLE INFO

Article history:
Received 15 March 2015
Received in revised form
28 May 2015
Accepted 7 July 2015
Available online 17 December 2015

Keywords:
Articulate body
Big data
High-resolution Image
Human recognition

ABSTRACT

Building recognition is a crucial and still challenging problem in computer vision and pattern recognition field, especially in human articulate body recognition. In this paper, we present a novel formulation for detecting articulated human bodies from high-resolution monitoring images. Firstly, we use a coarse detection process to detect the human object patches from monitoring images. To build a descriptor which contain color frequency information based on histograms of oriented gradient method. Then, a linear support vector machine was used to speed up detecting large image patches which contain the object joints. Secondly, towards an efficient detection, a refinement classification method is proposed to determine the patches that actually contain objects. Finally, to increase the size of image patches in order that the whole human objects can be included, an accelerated and improved salient mask is used to improve the performance of the dense scale-invariant feature transform descriptor. Experimental results on three datasets demonstrated our method effectiveness.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Building recognition is an important and useful technical in robot navigation, image retrieval, video surveillance field [2]. However, the method is still challenging in the accuracy and efficiency of building recognition successfully, especially for human body detection, since the human character model is an articulate object and the skeleton of human has many movable joints and motion parameters [5,6]. It is difficult to detect the articulated rigid bodies in high-resolution images [7], due to the dynamic backgrounds, occlusions and variations in lighting conditions and the number of building images is gigantic, which makes the training models to carry out a robust and fast building recognition model learning process[2].

In the recent years, many graph-based models are applied in multimedia and computer vision for building recognition. They can be used as geometric image descriptors[1] to enhance image categorization. For instance, these methods can be used as image high-order potential descriptors of superpixels [25–27,30]. The graph-based descriptors can be used as general image esthetic descriptors to improve image esthetics ranking, segmentation [22], photo retargeting and cropping [21,28–32].

The building recognition approaches can be briefly be divided into three categories: global feature-based building recognition, local feature-based building recognition, and local–global feature-based building recognition. Most of the approaches are concerned about detecting the interesting parts of the image. The Global feature-based approaches represent each building by a feature vector which is used as input for conventional classifiers, such as Support Vector Machine (SVM) [18]. The main weakness of global features is that they are sensitive to large changes in building images due to occlusions, clutters, lighting conditions, etc. [3,4]. The Local Feature-based approaches represent each building by a collection of local descriptors extracted at interest points. The method is casted as local descriptors matching. Zhang et al. used localized color histogram to retrieve a small number of candidate matching from the building data set [8,9]. Then the building recognition is refined by the matching descriptors associated with the local regions. But such an approach is very sensitive to large view changes of the same building. Chung et al. [10] proposed a sketch-based building recognition by defining a new local descriptor named multi-scale maximally stable external regions (MSER). The MSERs in the image are organized together into a sketch for matching between buildings. But the aforementioned local descriptor-based approaches are highly time-consuming due to the brute-force matching operations, which often face the obstacle to handle large-scale data sets. Local–global feature-based approaches utilize both global spatial information and local descriptors to represent each building image. Lazebnik et al. [11] developed spatial pyramid matching (SPM) by partitioning an image into increasing fine grids and computing histograms of local features inside each grid cell. However, the SPM method only shows the good performance when working together with a nonlinear SVM.

http://dx.doi.org/10.1016/j.neucom.2015.07.123
0925-2312/© 2015 Elsevier B.V. All rights reserved.

Furthermore, most of the recognition image methods cannot be directly applied to high-resolution image because they require expensive computation [12,13]. Many researchers aimed at the intelligent image processing [17,19,20]. Dalal and Triggs [14] proposed the histogram of oriented gradient (HOG) descriptor to human articulate body detection; the method is strongly based on the popular SIFT algorithm [15,16,23]. The HOG extracts the gradient direction histogram of the local area and considers it a feature of the local image. We present a method in this paper which can rapidly detect all objects that are similar to a goal object using the HOG features with color frequencies; our method concatenates the selected frequency of the RGB channels to improve the HOG descriptor. Then it produces an accurate classification to determine the goal object by applying a dense scale-invariant feature transform (SIFT) descriptor with a saliency mask to reduce background interference. And combined with a saliency measure, we use an adaptive optimization algorithm for determining the feature map, which better retains the contours of salient objects and reduces the computational time.

Our formulation that is based on the prior work of Xun at al. [7] mainly consists of two phases: Coarse Detection and Refinement Classification. The Coarse Detection phase frees us from high-resolution images and focuses the technique on small image patches of human bodies. The Refinement Classification accurately determines which patches are desired human articulate objects.

2. Algorithms

The process of our human articulated body recognition method is shown in procedure scheme (Fig. 1). There are two main phases in the process flow diagram: First, a coarse detection process in high-resolution images aims to find human objects with HOG features that are similar to the desired objects. Since the detection typically yields several small patches and does not include the entire human body, we expand the outline based on part models. Second, to extract these image patches and generate their dense SIFT features to determine whether they are objects that we desired, we are using a multi-detector and a multi-classifier in the coarse detection stage and the refinement phase separately. Finally, to mark the human objects in the original images and categorize the images according to the presence of the human body.

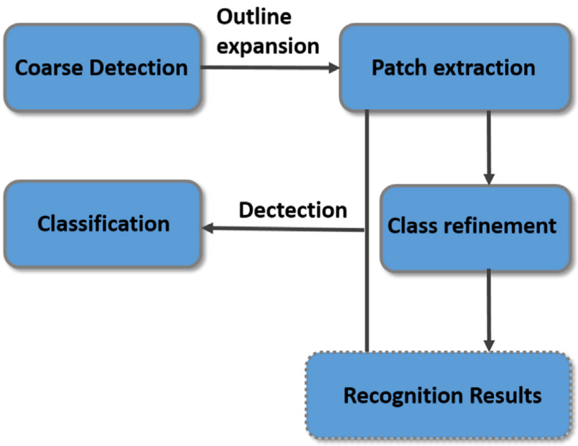


Fig. 1. Human articulated body recognition method process flow diagram.

2.1. Coarse detection process

The histograms of oriented gradient (HOG) descriptor are concatenated with all channels of the color space, such as CIELab, RGB, and HSV which are emphatically considered by our method. The process of coarse detection is importantly associated with the color frequencies, the length of the features and the computational complexity increased linearly with the number of color channels. The advantage for detection of human articulated body algorithm is that the most same-category objects part of human body can be clustered according to color. We call the color information of the objects "color frequency features". We concatenate the color frequency features with the HOG descriptors to improve the detection accuracy.

The purpose of coarse detection is to achieve a high detection rate in a small amount of time. In the coarse detection phase, we compute the horizontal and vertical gradients (,) of the targeting images in three dimensions by applying the filter $[-1\ 0\ 1]$. i presents the ith-dimensional color space. So we get the maximum gradient of the 3D at the same pixel by using

$$M(x,y) = \max_i \sqrt{\mathrm{GH}i(x,y)^2 + \mathrm{GV}i(x,y)^2}, \quad i = 1,2,3. \tag{1}$$

The dimension that contains the maximum value and the orientation and norm of the gradient are computed by using

$$\theta(x,y) = arctan\frac{G_H(x,y)}{\mathrm{G_V}(x,y)} \tag{2}$$

The HOG descriptors focus on the rigid components since there may be frequent changes to the relative positions of components. If we consider the linear SVM as a baseline classifier, our improved descriptor has a higher detection rate. There are so many image regions obtained by using the descriptors and the linear SVM detector, which may or may not contain the desired object components. So we expand the regions so that they contain the entire human objects to produce a precise classification.

By using a star-structured part-based model [24] shown in Fig. 2, we define the rigid components of the human objects as the goals of coarse detection and 'root' filter models. Then the outline of each goal part increases according to the possible relative position of the rigid components as following steps [7]:

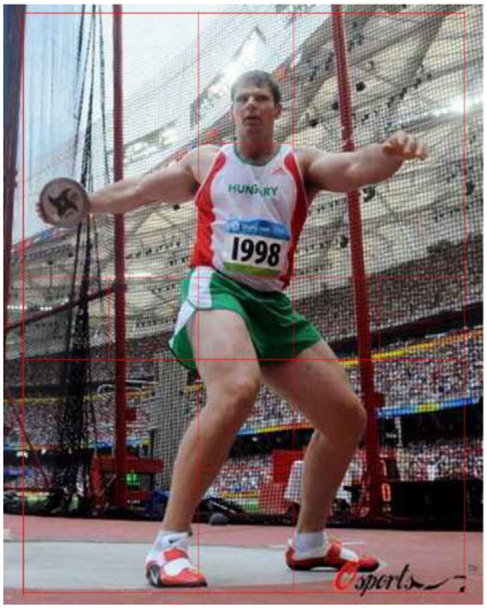


Fig. 2. Star-structured part-based model.

Download English Version:

https://daneshyari.com/en/article/405915

Download Persian Version:

https://daneshyari.com/article/405915

Daneshyari.com