Neurocomputing 119 (2013) 425–433

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

A two-layer framework for appearance based recognition using spatial and discriminant influences

Qi Li [a,*], Chang-Tien Lu [b]

[a] Department of Computer Science, Western Kentucky University, United States
[b] Department of Computer Science, Virginia Tech, United States

ARTICLE INFO

Article history:
Received 16 December 2010
Received in revised form
10 March 2013
Accepted 12 March 2013
Communicated by Zhu Xingquan
Available online 2 May 2013

Keywords:
Feature points
Fisher score
Linear discriminant analysis
Locality

ABSTRACT

Appearance of objects lie in high-dimensional spaces. Feature selection improves not only the efficiency of object recognition but also the recognition accuracy. In this paper, we propose a two-layer learning framework of feature selection using spatial and discriminant influences. The first layer selects a number of feature points of highest integrated influences by integrating spatial and discriminant influences, and the second layer refines the selection in terms of the discriminancy of these feature points measured by orientation histograms of their local appearances. The proposed framework can be categorized as a global appearance based recognition approach. Unlike popular projection methods, such as PCA, LDA, the proposed framework can present visual interpretability of selected features, which is desirable in bioinformatics and medicine informatics. We present two case studies: (i) embryo stage recognition and (ii) face recognition. Our case studies show the effectiveness of the proposed framework.

1. Introduction

Appearances of objects lie in high-dimensional spaces. For a given recognition task, feature selection aims to select most effective features (specifically, feature points) in order to reduce the computational cost of recognition and improve recognition accuracies. Features can be selected based on their spatial influences [15,26,31,53], i.e., the bottom-up scheme [29,28,48,59]. For example, the Harris detector [15] uses gradient auto-correlation of image points to define their spatial influences. The bottom-up scheme aims to output feature points repeatable across different imaging conditions, which helps construct robust and compact representation of image data. The bottom-up scheme has a wide range of applications, such as object recognition [18], image retrieval [33]. The bottom-up feature selection is an important step to build a generative model for object recognition [54,8]. A generative model is basically a graph model with a relatively small number of features [18]. Generative models are strong in addressing "weak-alignment" recognition tasks where the shapes of different objects contain significant variations. Features can also be selected in terms of class or context information, i.e., top-down schemes[13,29,28,48,59]. Gao and Vasconcelos [13] argued that spatial information (such as edge, corners) may not always reveal good saliency of visual objects, and thus proposed a discriminant top-down selection method for visual recognition, where the discriminancy is determined by the maximum marginal diversity [49].

Recently, the integration of the bottom-up and top-down feature selection received extensive attention in the area of visual classification [29,28,48,59], including object detection [34], object recognition [18], and scene understanding [48]. For example, Holub and Perona [18] proposed a model to combine the generative model and Fisher kernels, which brings considerable improvement of the performance of generative models. To speedup object detection, Navalpakkam and Itti [34] proposed a model to integrate bottom-up and top-down attention, where the top-down component uses accumulated statistical knowledge of the visual features of the desired search target and background clutter, to optimally tune the bottom-up maps such that target detection speed is maximized. More related work will be presented in Section 2.

In this paper, we propose a two-layer learning framework for appearance based recognition via a hierarchical usage of spatial and discriminant influences. The proposed framework assume that images (objects) are aligned so that image points at the same location in different images have "correspondence", e.g., their intensities tend to be correlated to each other. In other words, the proposed framework stands on the techniques of image registration [4,17,60], object localization [3,52,11], and image segmentation [55,30,35].

The main idea of the proposed framework is illustrated in Fig. 1. Given a set of training images, the first layer aims to select a

* Corresponding author. Tel.: +1 270 7456225.
E-mail addresses: qi.li@wku.edu (Q. Li), ctlu@vt.edu (C.-T. Lu).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.03.015

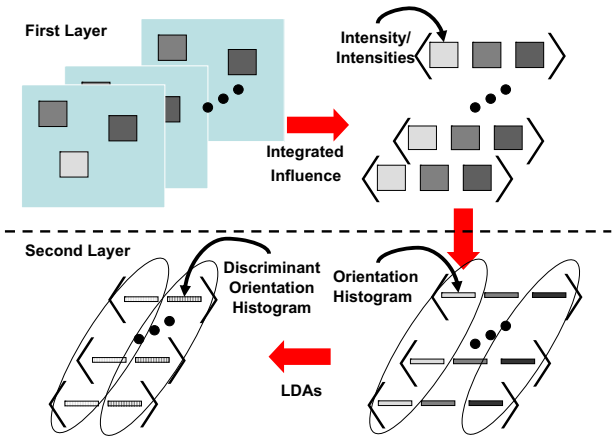

Fig. 1. Two-layer learning framework for appearance based recognition. The first layer selects a number of feature points of maximal integrated influences, and the second layer applies Linear Discriminant Analysis to an ensemble of descriptors of these feature points (constructed by orientation histogram) and obtains an ensemble of most discriminant representations of orientation histograms.

number of feature points (visualized as small blocks in Fig. 1) by integrating spatial and discriminant influences. The major output of the first layer is the locations of feature points. In the second layer, we first construct the orientation histogram (visualized by a row vector in Fig. 1) for the local appearance of each feature point (of each image). The collection of orientation histograms of the same feature points of all training images forms an ensemble of instances (visualized by an ellipse). Linear Discriminant Analysis (LDA) [9,45,1] applies to an ensemble of instances to estimate the discriminancy and compute the discriminant representation of orientation histograms. So the major output of the second layer of the proposed framework is a reduced set of feature points and LDA subspaces associated with the highest discriminancy score.

The rationale of the hierarchical design of the proposed framework lies in the following comparison between intensity blocks and orientation histograms:

- *Efficiency*. Constructing orientation histograms is much more computationally expensive than constructing intensity blocks, which is the rationale of introducing orientation histogram in the second layer rather than the first layer, i.e., being applied to selected feature points rather than all image points.
- *Sensitivity to localization*. Orientation histograms are less sensitive to localization error than intensity blocks, which is the rationale of introducing Linear Discriminant Analysis (LDA) to estimate the discriminancy of orientation histograms. Note that LDA assumes that data fits Gaussian distribution.

In experiments, we present two case studies to demonstrate the effectiveness of the proposed framework. The first case study is on the recognition of developmental stages of Drosophila embryos based on gene expression pattern images [22,14]. The role of Drosophila (fruit fly) in explicating the function and interconnection of animal genes has established the species as a major model organism [44]. In situ hybridization is a recent technique to document gene expression pattern of embryos along their different developmental periods [7]. (An expression region indicates the response of a gene to a probe RNA.) These documents, a set of embryos images contain rich information on the spatio-temporal patterns that are extremely valuable for the study of gene–gene interaction. Dark regions in an embryo image indicate expression patterns of genes. Given two standardized images of embryos (of pixel-to-pixel correspondence) at the same developmental stage, the interaction strength of two genes can be quantified by computing the similarity of expression patterns [23,20,19,38,27,10], e.g., the ratio of overlapping expression regions of the images. Compared with in situ hybridization, the widely used microarray technique reveals very limited spatial pattern information. The gene expressions obtained from microarray images are, precisely speaking, the average expression levels. The second case study is on face recognition where we use dataset PIE [43] that has 68 human faces. These studies convince us the effectiveness of (i) the integrated influences in selecting good feature points, and (ii) the discriminant representation of orientation histograms.

The rest of the paper is organized as follows: In Section 2, we present related work. In Section 3, we introduce locality oriented Fisher discriminant scores. In Section 4, we propose an integrated model. In Section 5, we propose discriminant representation of orientation histograms. Two case studies are presented in Section 6, and conclusions and future work are given in Section 7.

2. Related work

Appearance based recognition can be roughly categorized into two different strategies: (i) global approaches and (ii) local approaches [40]. Global appearance based recognition usually assumes object regions have been aligned. It has been successfully applied to some weakly textured images, such as face images [56]. A global approach is popularly performed in terms of a projection method, such as Principal Component Analysis (PCA) [21], Linear Discriminant Analysis [1], and their high-dimensional variations: Generalized PCA [51], tensor Discriminant Analysis [46], etc. A global appearance can be effectively "encoded" into a very low dimensional vector for recognition, and thus brings appealing recognition efficiency. Besides the advantage of recognition efficiency, projection methods can achieve satisfying recognition accuracy if global appearances do not contain significant local outlier appearances. A limitation of projection methods is that their performance is difficult to interpret. Note that interpretability is desirable in many applications, such as object categorization [39] and bioinformatics [16]. It is worth noting that sparse learning, as a projection based feature extraction scheme, recently received attention [58,24,47,57] since features extracted by a sparse learning method can be interpreted psychologically and physiologically [58].

In contrast to global approaches, local approaches are robust with respect to localization error and local outlier appearances [41,26]. (Precisely speaking, local approaches do not require object localization.) In a local approach, a set of repeatable/stable image point/regions are first extracted by an interest point/region detector [42,26,31], and then distinctive descriptors are constructed to represent an image. However, local approaches are computationally expensive. Moreover, local approaches are not effective for weakly textured objects, such as face images, feature selection method instead of a projection method.

Walther et al. [53] proposed a bottom-up model for selective attention, where bottom-up saliency map is contributed by the color feature maps, intensity feature maps, and orientation feature maps. They showed that the proposed bottom-up visual attention can strongly improve learning and recognition performance in the presence of large amounts of clutter.

Vasconcelos [49] proposed a discriminant feature selection via maximization of marginal diversity (MMD); for multi-class problems, one-versus-all strategy is applied. Vasconcelos and Vasconcelos [50] proposed an information theoretic feature selection to achieve a good balance between maximizing the discriminant power of selected (local) features and minimizing their redundancy. The method is tested on image retrieval, where the