Pattern Recognition Letters 54 (2015) 82–88

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A new shape prior model with rotation invariance ☆

Chunhui Zhang, Jimin Liang*, Jun Zhang, Heng Zhao

School of Life Science and Technology, Xidian University, Xi'an, Shaanxi 710071, China

ARTICLE INFO

Article history:
Received 22 February 2014
Available online 8 January 2015

Keywords:
Shape prior model
Rotation invariance
Distance transforms

ABSTRACT

Object detection methods based on keypoint localization are widely used, and the rotation invariance is one of the fundamental issues to consider. This paper proposes a novel shape prior model with rotation invariance. The proposed shape prior model discards all orientation-involved features and only uses the distance features among keypoints, hence it is competent to detect objects with a rotation of the arbitrary angle when combined with local appearance description with rotation invariance. In the stage of detection, belief propagation algorithm is employed, so that our method no longer needs the initial position of the keypoints. Furthermore, we generalize the classical distance transforms, the generalized distance transforms make the beliefs to be calculated in a nearly linear time. Experiments were carried out on face category and touring-bike category in the Caltech-256 database. The results demonstrated that the proposed method achieved a strong robustness of rotation.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The object detection methods based on keypoint localization have attracted increasing attention [1–11]. They can achieve an accurate location for all of the keypoints at once, because they combine the appearance information around each keypoint and the information about spatial relationships among them. Thus, methods of this kind need one appearance feature to describe the local appearance information, and one shape prior model to describe the spatial information among the keypoints. In detection, consideration must be given to the matching of both the local appearance and the spatial relationships simultaneously.

Generally, the shape prior affects the performance of the whole system directly. In many circumstances, rotation invariance is an important property required for the shape prior model. In this paper, two issues about the shape prior are discussed: (1) the way of the shape description; (2) the shape detection (optimizing the objective function).

The influence of the first issue on the keypoint localization is huge. Some kinds of shape prior models [5,7–9] use the orientation-involved features to describe the shape. For example, the shape prior of pictorial structures model (SPS) [5,7] employs the tree to represent the frame of the shape prior, where each keypoint is a node of the tree. The edge represents the spatial relationship between two keypoints, and the coordinate difference of these two keypoints is treated as the feature of this edge. However, it is difficult to match these shape priors to the rotated shape due to their high sensitivity to rotation. To obtain a shape description with rotation invariance, there are mainly two ideas to follow: (1) estimating the orientation of the object and transforming the current object into the upright object and finally re-matching the upright object with the model; (2) using the orientation-insensitive shape description. Most shape prior models [10–13] belong to the first category, which possess the rotation invariance by estimating a rotation matrix. In the training stage of these methods, the training shapes are obtained by being manually labeled, and then usually aligned with an upright regular shape. After that, the components of interest are extracted from the aligned shapes. For example, the components of interest in the sparse shape prior [12,13] are the sparse code-books, while the components of interest in ASM (active shape model) [10] or AAM (active appearance model) [11] are the principal components. The components of interest are combined linearly to form the upright version of the object shape, and finally the upright shape is rotated by the rotation matrix to obtain the practical shape. From the above description, the procedure of training is verbose, and it is difficult to estimate an accurate rotation matrix.

Another issue for keypoint localization is shape detection. Some methods (e.g. ASM [10], AAM [11] or sparse shape prior [12,13]) adopted the alternating policy during detection. As a result, the detection usually becomes two alternating procedures: one procedure that matches the components-of-interest-combined shape with the upright current shape transformed from the current estimated shape, another procedure that updates the rotation matrix via

☆ This paper has been recommended for acceptance by G. Moser.

* Corresponding author: Tel.: +86 29 81891060, +86 29 81891070; fax: +86 29 81891060.
E-mail address: jiminliang@gmail.com, jimleung@mail.xidian.edu.cn (J. Liang).

http://dx.doi.org/10.1016/j.patrec.2014.12.008
0167-8655/© 2015 Elsevier B.V. All rights reserved.

procrustes analysis [14]. However, this kind of the detection framework implies that a proper initial localization is necessary [11] and that the radii of convergence are small. Actually, SPS [5,7] employed belief propagation (BP) algorithm [15] during detection. BP provides a new idea for the detection, in which the initial value is no longer necessary. In the BP algorithm, a kind of information called the belief is exchanged between the keypoints. After multiple iteration, each keypoint would gather the information of all the other ones, allowing the location to be more accurate. More attention [16–18] has been paid to BP since the appearance of fast distance transforms [18]. However, to the best of our knowledge, the current BP based methods do not consider the rotation of the object, so the fast distance transforms are not applicable in our case.

This paper proposes a new shape priori model which uses the orientation-insensitive shape description directly. The proposed shape prior can be combined with some rotation-invariant local appearance description [19] to achieve the detection with arbitrary rotation. We need to construct a graphical model to represent the frame of the shape prior. As same as SPS, each edge shows the spatial relationship between two keypoints. In order to remove the influence of objective rotation on detection, only the distance is treated as the feature of the edge. Since dynamic programming algorithm used in [5,7–9] does not allow cycles in the graphical model [20], we adopted the BP algorithm to obtain the accurate detection result. However, due to the consideration of the rotation, the form of the beliefs (called the generalized beliefs) used in our method is different from that used in [5,8]. So we generalized the distance transforms algorithm [18] to make the calculation of the generalized beliefs be accomplished in nearly linear time.

2. Methods

The major function addressed by our method can be summarized as: given a testing image in which the orientation of the object is arbitrary, output the positions of all of the keypoints.

2.1. Shape prior

In this section, the shape prior model is built using the training set, where the shape in each training image is labeled manually. The ith shape is described as the matrix $d_i \in R^{n\times 2}$, where n is the number of keypoints. Each row of d_i is the coordinate of one keypoint. The set of training shape can be denoted as $D = \{d_1, d_2, \ldots, d_k\}$. Let $L \in R^{n\times 2}$ be the shape of the object in the testing image. The proposed prior model of the shape L is defined as

$$F(L) = \sum_{(v^i,v^j)\in E} \left(\sqrt[2]{(y_i - y_j)^2 + (x_i - x_j)^2} - m_{ij} \right)^2, \qquad (1)$$

where $V = \{v^1, v^2, \ldots, v^n\}$ represents the set of keypoints, and $[y_i, x_i]$ denotes the coordinate of keypoint v^i. m_{ij} is the mean distance between v^i and v^j, which can be estimated by training set D. V is the set of vertices of G, and E is the set of edges of G. In our experiment, if the number of keypoints was larger than 2, we constructed the minimum Hamiltonian cycle taking the variance of the distance as the weight, so the constructed minimum Hamiltonian cycle was treated as the graph. The aim is to reduce the distance volatility between the nearby keypoints, and raise the detection precision. In addition, if the spatial constraint of some keypoint is weak, one additional edge can be optionally added to be connected to that keypoint.

As can be seen in Section 2.2, Eq. (1) imposes penalty to the deviation of some distances from the mean distances. In other words, it restrains the distances between the keypoints around their mean values. In addition, involving the distances only, our shape prior has completely removed the influence of objective rotation on detection.

2.2. Shape detection

Given a testing image with the image size of $H \times W$, we search the optimal shape L^* by minimizing the following object function:

$$\begin{aligned} P(L) &= \lambda F(L) + R(L) \\ &= \lambda \sum_{(v^i,v^j)\in E} \left(\sqrt[2]{(y_i - y_j)^2 + (x_i - x_j)^2} - m_{ij} \right)^2 \\ &\quad + \sum_{v^h\in V} r_h(y_h, x_h), \end{aligned} \qquad (2)$$

where $r_h(y, x)$ represents the cost value when the hth keypoint is placed at $[y, x]$. In our paper, it is determined by inputting the local appearance feature around the keypoint into SVM classifier [21,22]; the details can be found in Section 3. Minimizing Eq. (2) ensures not only the low cost value of each keypoint, but more importantly, also the intrinsic spatial relationship between keypoints. λ is the weight of shape prior model, which exhibits the degree of elasticity within the spatial relationship. If λ is excessively large, the shape prior would be solidified and lose flexibility. Because of the existence of the cycle in our graphical model, we adopt the BP algorithm [15] to optimize Eq. (2). The pseudo-code of the BP algorithm is shown as follows:

Algorithm 1 BP algorithm.

Initialization:

Set $t = 0$;

Set q_{ij}^0 =Zero matrix, q_{ji}^0 =Zero matrix;

Iteration:

while the algorithm is not convergent **do**

  for all $(v^i, v^j) \in E$ **do**

    for all $[y_j, x_j] \in \{1, \ldots, H\} \times \{1, \ldots, W\}$ **do**

$$\begin{aligned} q_{ij}^{t+1}(y_j, x_j) = \min_{y_i,x_i} \Bigg\{ &\left(\sqrt[2]{(y_i - y_j)^2 + (x_i - x_j)^2} - m_{ij} \right)^2 \\ &+ r_i(y_i, x_i)/\lambda + \sum_{s\in(\zeta(i)-j)} q_{si}^t(y_i, x_i) \Bigg\}; \end{aligned} \qquad (3)$$

    end for

    for all $[y_i, x_i] \in \{1, \ldots, H\} \times \{1, \ldots, W\}$ **do**

$$\begin{aligned} q_{ji}^{t+1}(y_i, x_i) = \min_{y_j,x_j} \Bigg\{ &\left(\sqrt[2]{(y_j - y_i)^2 + (x_j - x_i)^2} - m_{ji} \right)^2 \\ &+ r_j(y_j, x_j)/\lambda + \sum_{s\in(\zeta(j)-i)} q_{sj}^t(y_j, x_j) \Bigg\}; \end{aligned}$$

    end for

  end for

  t=t+1;

end while

for all $j \in \{1, \ldots, n\}$ **do**

  $L^*(j, :) = \arg\min_{y_j,x_j} \{ r_j(y_j, x_j) + \sum_{s\in\zeta(j)} q_{sj}^t(y_j, x_j) \}$;

end for

Output:

The position of all the keypoints L^*.

where the matrix q_{ij}^t of $H \times W$ denotes the belief spreading from v^i to v^j after the tth iteration, $\zeta(i)$ denotes the set of vertexes adjacent to v^i in G. Intuitively, the calculation of belief needs a two-hierarchy loop. One loop is implemented for $[y_j, x_j]$, and another loop is implemented for $[y_i, x_i]$, so the time complexity is $O((H \times W)^2)$. However, Pedro F. Felzenszwalb reduced the time complexity of the distance transforms (shown as Eq. (4)) to $O(H \times W)$ in [18].

$$D_f(p, k) = \min_{y,x} \{ (y - p)^2 + (x - k)^2 + f(y, x) \}, \qquad (4)$$

Download English Version:

https://daneshyari.com/en/article/536306

Download Persian Version:

https://daneshyari.com/article/536306

Daneshyari.com