Int. J. Electron. Commun. (AEÜ) 69 (2015) 1272–1281

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

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

A novel multi-scale and multi-expert edge detector based on common vector approach

Kemal Özkan *, Şahin Işık

Department of Computer Engineering, Eskişehir Osmangazi University, Meselik Campus, 26489 Eskişehir, Turkey

ARTICLE INFO

Article history:
Received 13 November 2014
Accepted 13 May 2015

Keywords:
Multi scale
Multi expert
Edge detection
Common vector approach
Non-maximum suppression

ABSTRACT

In this paper, an information fusion algorithm – MSME-CVAED – is proposed for multi-scale and multi-expert edge detection. Well-known operators, called experts, have been applied to distinct scales derived by smoothing the gray image with Gaussian functions having different variance values. Common characteristics of edge points are processed to merge the information obtained from each scale, based on the concept of common vector approach. Once a single gradient map obtained, a smart non-maximum suppression operation is carried out to obtain a binary edge map. Later, an edge segment validation process is introduced based on Helmholtz principle, a common method in which edge segment validation is carried out with the "*a contrario*" approach using the number of false alarms concept. Experiments on popular datasets of ROC curves and RUG show that the proposed method achieves superior results in terms of *F-measure* and Mathews correlation coefficient score, compared with some recently published edge detectors.

1. Introduction

Multi-scale representation of images is a formal theory of handling images at different scales. It is inspired by the abilities of the human vision system and aims to be able to distinguish significant structures across a range of scales with different amounts of information in their scenes. To this end, a vast amount of image transform filters have been developed to decompose images using scale-space to extract features, suppress noise and perform smoothing. Unfortunately, the multi-scale approach involves certain problems, including the proper selection of scale for the scene and the significant way for handling the scene. Nevertheless, a good edge detector should make a trade-off between well-localized, continuous and thin edges that match the real edge structure of processed image, and jittered, purely qualified and noisy edges. In this respect, multi-scale edge detection is a specific solution developed to overcome such problems by performing single edge detection at multiple scales or merging edge cues or information from distinct scales.

Traditionally, images are smoothed at several scales in multi-scale edge detection methods [1,2]. The idea of using the multi-scale approach for edge detection was first explicitly proposed by Witkin [3] in which a Gaussian filter is applied to the various scales to reveal the zero-crossings of the second order derivative of a 1-D signal. With this work, Witkin paved the way for modeling the edge as a function of scale, thus inspiring many studies. Bergholm proposed a multi-scale edge detection method by running the Canny operator at decreasing scales and combining edges by considering two consecutive edge maps obtained from different scales based on the gradient of Gaussian [4]. Additionally, the analysis and extraction of different types of edges presented in the study of Lu and Jain [5,6], based on the behavior of zero-crossings, is described with 35 rules to obtain well-localized edges. Recently, some new approaches have been proposed, including a multi-scale method for edge detection based on increasing Gaussian smoothing and edge tracking [7], a statistical approach to combining edge cues at multiple scales using data driven probability distributions [8] and an adaptive threshold edge detection algorithm based on dyadic wavelet transform by multi-scale analysis [9]. Moreover, some techniques for multi-scale edge detection based on wavelet transform have been developed, such as a wavelet domain vector hidden Markov tree (WD-VHMT), which models the statistical properties of multi-scale and multidirectional (subband) wavelet coefficients of an image [10] and a multi-scale edge detection algorithm based on wavelet transform that is concentrated on some classical gradient operators [11].

The reason for using a multi-scale scheme in edge detection is that the strategy is associated with human visual system and the first principles of physics of vision. In fact, we have to look from different perspectives and consider various possibilities to make a

* Corresponding author. Tel.: +90 2222393750.
E-mail addresses: kozkan@ogu.edu.tr (K. Özkan), sahini@ogu.edu.tr (Ş. Işık).

http://dx.doi.org/10.1016/j.aeue.2015.05.011
1434-8411/

proper decision about an event in real life. In Ref. [7], techniques on multi-scale edge detection have been classified into three families based on the way the relevant information is gathered at different scales. According to the study, the first group detects the edges at different scales and processes them to make use of the combined results to yield a single edge map, e.g., Bergholm's method [4]. The second group illustrates a different strategy, in which the cues and information for edge points are derived and later merged, with the information obtained from each scale, e.g., Refs. [12,13]. The third group aims to determine a single, non-homogeneous scale that can be used at each pixel or sub region by observing the local characteristics, e.g., Refs. [14,15]. Our proposed method falls in the second group because gradient maps are computed at each scale, obtained with a Gaussian filter of varying variance size, after which a single gradient map is constructed from these gradients based on the concept of common vector approach (CVA), and at the end of which, a smart edge extraction procedure is performed to give an edge map that has good quality and performance. Each edge detection operator is regarded as an expert. As mentioned above, the most widely known edge detection operators are employed for the sake of multi expert edge detection operations. Each expert is used at three Gaussian scales. The main difference of this study from others can be attributed to CVA, which is applied to the gradient maps to fuse the information of the computed gradients and obtain a single detection result.

2. Overview of the common vector approach

CVA is a subspace-based recognition method that has been used in many pattern recognition tasks, e.g., voice [16] and face recognition [17], spam e-mail classification [18], image denosing [34]. A common vector is what is left when the differences between training vectors are removed from class members, and the common vector is invariant throughout the class. In CVA, the number of vectors is either sufficient or insufficient in cases where the vector dimension (n) is lower or higher than the number of vectors (m) available, respectively. In the case of insufficient data, where the number of vectors is less than or equal to the dimension of a vector ($m \leq n$), the common vector for a class can be obtained by using the Gram–Schmidt orthonormalization or subspace techniques. For this work, a 2-D 3×3 reference block has been transformed into a 1-D vector of 1×9. Additionally, by using three scales and three edge detector experts, the number of vectors (m) becomes 9. If they are combined to form a 9×9 matrix, then the dimension of vector (n) is the equal to 9. This is obviously a case of insufficient data because the number of vectors is equal or less than the dimension of a vector ($9 \leq 9$). If the vector of gradient map obtained from each edge detector expert at any scale is represented as $\vec{a}_i$, then the training set is constituted as $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_9 \ \in \Re^{MxN}$. Also, the covariance matrix Φ of the training set is obtained from Eq. (1)

$$\Phi = \sum_{i=1}^{9} \left(\vec{a}_i - \vec{a}_{\text{ave}}\right)\left(\vec{a}_i - \vec{a}_{\text{ave}}\right)^T \quad (1)$$

In Eq. (1), $\vec{a}_{ave}$ indicates the average vector. In CVA, the feature subspace is separated into two perpendicular subspaces: difference subspace P and indifference subspace $P^{\perp}$. In the case of insufficient data, the covariance matrix (Φ) will have $n-m+1$ zero eigenvectors [16]. Also, P and $P^{\perp}$ are spanned by the eigenvectors corresponding to nonzero and zero eigenvalues, respectively. When the eigenvectors are ordered from largest to smallest as $\lambda_n \geq \lambda_{n-1} \geq \cdots \geq \lambda_{n-m+2} \geq \lambda_{n-m+1} \geq \cdots \geq \lambda_2 \geq \lambda_1$, the smallest $n-m+1$ eigenvalues are all zero. That is,

$$P = \text{span}(u_i), \lambda_i \neq 0, \quad i = n, n-1, \ldots, n-m+2, \quad \text{difference subspace}$$

$$P^{\perp} = \text{span}(u_i), \lambda_i = 0, \quad i = n-m+1, n-m, \ldots, 1, \quad \text{indifferent subspace}$$

A given vector can be represented as $\vec{a}_i = \vec{a}_{\text{com}} + \vec{a}_{\text{diff}}$ where the common vector, $\vec{a}_{\text{com}}$, for a class, can be obtained by a projection of any feature vector onto the indifference subspace ($\vec{a}_{\text{com}} = P^{\perp}\vec{a}_i, \quad \forall i = 1, 2, \ldots, m$) and the difference vector, $\vec{a}_{\text{diff}}$, for a class, can be obtained by a projection of any feature vector onto the difference subspace ($\vec{a}_{\text{diff}} = P\vec{a}_i$).

3. Edge detection procedure using CVA

The fusion of information and cues from the outputs of distinct operators at each scale is another important procedure in the literature of edge detection. Generally, the edge combination process involves labeling similar edges based on the minimum error criterion to form a single edge map. Although the edge labeling procedure is an easy way to combine edges in the case where the characterization of edge behavior is similar at all scales, edge labeling is not convenient for tracking the edges in a fine-to-coarse scheme because edges that disappear never reappear when going from a fine to a coarse scale. The edge fusion procedure of proposed method is carried out by using CVA, as demonstrated in Fig. 1. Since CVA exposes the common properties of a class, it is expected that a common gradient map can be obtained based on the concept of CVA. Therefore, it is employed on the vectors obtained from the 2-D gradients of aforementioned operators. The common edge cues and information situated at each scale is combined into a single gradient map. Later, this process is followed by non-maximum suppression, dual thresholding and edge segment validation steps. Main steps of the proposed edge detection procedure using CVA are given below.

Algorithm: Edge detection procedure using CVA1.
Calculating 9 (obtained from 3 scales and 3 experts) distinct gradient maps
2. A 3×3 region is taken from each gradient map and converted into a 1×9 vector
3. All vectors are combined to obtain a single 9×9 matrix (because there are 9 gradient maps)
4. CVA is applied to the 9×9 matrix size by considering the insufficient data case where the dimension vector (9) is equal to the size of vectors (9). Then, a 9×1 common vector is obtained from the processed matrix
5. Later, this common vector is reshaped as a 3×3 matrix.
6. Finally, these common matrices are combined to form a single MxN gradient map by using weighting average to reduce the border effects.
7. Apply non-maximum suppression and thresholding process which was proposed in [27]
8. Apply edge segment validation process which was proposed in [19]

To explain the CVA based information fusion, a visual demonstration of stages involved in algorithm is given in Fig. 2. By looking steps in Fig. 2, we can see that a 3×3 frame is processed by focusing a region of the *Home* image. In given figure, labels from a1 to a9 refer to nine distinct gradient maps that computed after performing three experts (Canny, Sobel and Prewitt) on three scales.

To satisfy the insufficient data case, generated all 2-D 3×3 gradient frames are converted into vector format with 9×1 dimension and combined to construct a 9×9 matrix, where the dimension vectors is equal or less than the number of vectors ($9 \leq 9$). Specifically, a1 to a3, a4 to a6 and a7 to a9 indicate gradient vectors computed on 1st, 2nd and 3rd scale by performing Canny, Sobel and Prewitt filters, respectively. In next step, the mean of matrix is subtracted from all vectors to make data centralized, which is also known as mean removal process. Once the data converged into center, the eigenvalue and eigenvectors are computed based on the covariance matrix. While difference subspace indicates eigenvectors related to eigenvalues that is not equal to zero, oppositely, indifference subspace refers to eigenvectors corresponding to eigenvalues that equal zero. The common vector for this sample can be computed from the projection of any vector among a1 to a9 into the indifference subspace, which occurs in case of the insufficient data. The