IJDAR (2015) 18:137–151
DOI 10.1007/s10032-015-0238-y

SPECIAL ISSUE PAPER

A new ring radius transform-based thinning method for multi-oriented video characters

Yirui Wu · Palaiahnakote Shivakumara · Wang Wei · Tong Lu · Umapada Pal

Received: 6 June 2014 / Revised: 30 December 2014 / Accepted: 10 January 2015 / Published online: 24 January 2015

Abstract Thinning that preserves visual topology of characters in video is challenging in the field of document analysis and video text analysis due to low resolution and complex background. This paper proposes to explore ring radius transform (RRT) to generate a radius map from Canny edges of each input image to obtain its medial axis. A radius value contained in the radius map here is the nearest distance to the edge pixels on contours. For the radius map, the method proposes a novel idea for identifying medial axis (middle pixels between two strokes) for arbitrary orientations of the character. Iterative-maximal-growing is then proposed to connect missing medial axis pixels at junctions and intersections. Next, we perform histogram on color information of medial axes with clustering to eliminate false medial axis segments. The method finally restores the shape of the character through radius values of medial axis pixels for the purpose of recognition with the Google Open source OCR (Tesseract). The method has been tested on video, natural scene and handwritten characters from ICDAR 2013, SVT, arbitrary-oriented data from MSRA-TD500, multi-script character data and MPEG7 object data to evaluate its performances at thinning level as well as recognition level. Experimental results comparing with the state-of-the-art methods show that the proposed method is generic and outperforms the existing methods in terms of obtaining skeleton, preserving visual topology and recognition rate. The method is also robust to handle characters of arbitrary orientations.

Keywords Ring radius transform · Multi-oriented video characters · Medial axis · Thinning · Optical character recognition · Recognition

Y. Wu · W. Wei · T. Lu (✉)
National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
e-mail: lutong@nju.edu.cn

Y. Wu
e-mail: wuyirui1989@163.com

W. Wei
e-mail: sncweiwang@163.com

P. Shivakumara
Faculty of Computer Science, University of Malaya, Kuala Lumpur, Malaysia
e-mail: hudempsk@yahoo.com

U. Pal
Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India
e-mail: umapada@isical.ac.in

1 Introduction

Thinning is an important preprocessing or normalization step before feature extraction in many document analysis and computer vision tasks, such as fingerprint identification, biometric authentication using retinal images, signature verification, sketch matching and sketch-based image retrieval [1–7]. It is an integral part of character or object recognition methods because it is invariant to pixel level thickness of strokes [2–7]. Therefore, thinning is considered as an active research area for researchers.

Generally, a good character thinning algorithm must satisfy the following characteristics: (1) It should produce a thin skeleton of the input image, (2) it should preserve the shape or the visual topology of the character, and (3) it should be robust to orientation or font variations, noises, disconnections and distortions caused by blur, low resolution and complex background. So far, plenty of methods have been proposed in the past decades for thinning. However, we can see these methods rarely meet all the above criteria simultaneously. In

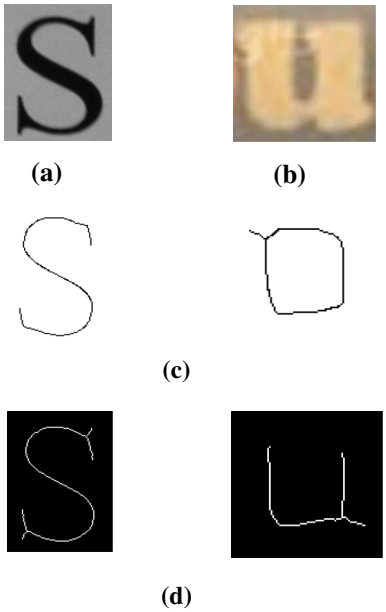


Fig. 1 Illustration for video character thinning in comparing with existing methods. **a** High resolution with plane, **b** video character image, **c** thinning output of (**a**) and (**b**) obtained from the existing method [1], **d** thinning output of the proposed method

particular, most of the methods focus on criterion (1) and (2) since they are usually developed for high-resolution images with plane backgrounds, but not for low resolution images obtained from video, where we can also expect different orientations and contrasts, complex backgrounds, font or font size variations, etc. [6,7]. Besides, video generally contains two types of texts: Caption text which is edited or superimposed, and scene text which exists naturally and is embedded on background. Since caption text is edited, it has clear visibility and clarity. But scene text is a part of an image, so its characteristics are unpredictable [8,9]. The presence of both the texts in video makes the problem more complex and challenging. For example, Fig. 1 illustrates the thinning results of the proposed method and another existing method [1] for the input images consisting of a high-resolution scanned character "S" and a low-resolution video character "u" as shown in Fig. 1a, b, respectively. It is observed from Fig. 1 that the existing thinning or skeleton method [1] does not work well for video characters as we can see the loss of the shape (character "u" looks like another character "O"), while it works well for the high-resolution character "S" as shown in Fig. 1c. On the other hand, the proposed method works well for both the characters as shown in Fig. 1d.

It is evident from the work [10–12] that conventional skeleton algorithms sometimes fail to preserve the shape of a character because they are sensitive to background and contrast. Since the methods focus on text detection or script identification but not recognition, skeleton does not affect much for the overall performances of the methods. In contrast to these tasks, a recognition task requires the complete shape and topology for the original character to achieve a good recognition rate. Therefore, there is a great demand for a robust thinning algorithm that works for video characters as well as scanned and camera-based characters by satisfying the three criteria discussed above.

2 Related work

As discussed in the introduction section, most of the methods in literature aim at thinning binary images scanned by a high- resolution scanner [1–7]. Therefore, thinning methods can be classified as sequential methods, parallel methods and medial axis methods. *Sequential methods* delete contour pixels iteratively in a predefined order, hence these methods are called non-isotropic. To overcome this problem, *parallel methods* have been developed. These methods delete contour pixels based on the results of previous iterations. Though these methods solve most of the problems of binary images, they are sensitive to noise and heuristics. On the other hand, *medial axis methods* produce medial or central lines of the pattern using distance transformation. These methods received considerable attention as they are insensitive to noise to some extent compared to the above two categories. However, since these methods use single pass for finding each central line, the methods give slightly distorted results at local regions, namely corners, junctions and intersections. Thus, they cannot well preserve the shape of the original character. The methods in [13,14] work for gray images and require a binarization step before performing thinning. Chen and Yu [15] proposed an entropy-based method for thinning noisy images. The method computes maximal information for a circular range on each pixel. The symmetry score of pixel distribution is used to obtain the skeleton of the image. However, the performance of this method decreases when noise increases. Hoffman and Wong [13] proposed a thinning method for both binary and gray images based on scale space filtering. The method finds peak, ridge and saddle points in gray-filtered images, which are called the most prominent ridge line pixels and used to get the skeleton. The quality of the pixels is increased via image scale space pyramid. However, the parameters have to be set by user manually. Cai [14] proposed a thinning method based on oriented Gaussian filters. The method is developed to solve the problem of contour noises that present in handwriting and finger prints caused by pen perturbations, scanning documents and images. This method classifies pixels into edges, valleys and ridges using Gaussian-oriented filters. The method uses intensity surfaces of ridge energy images and principal directions to get the skeleton of the image. However, this method focuses on thinning fingerprint and handwriting. From the above discussion, we can notice that none of the methods addresses the noise

problem properly. To overcome this problem, recently, Chatbri and Kameyama [1] proposed a method based on scale space filtering to make thinning algorithms robust to noises in sketch images. This method derives multiple representations of an input image with multiple scales of filtering. The filtering scale that gives the best tradeoff between noise removal and shape distortion is selected. Overall, the method proposes adaptive preprocessing in which various thinning algorithms are used to automatically estimate the optimal amount of filtering to deliver neat thinning results. However, the method is only limited to sketches and handwriting. Bag and Harit [16] proposed a method for solving the problem of spurious strokes of deformities because this problem is common for almost all the thinning methods. The method uses shape characteristics of text to get the skeleton that is nearly the same as the true character shape. It is not clear whether the method works well for scaled and different oriented characters since its objective is to thin printed/synthetic character images.

The above discussion reveals that the primary focus and scope of the existing methods is to thin scanned character images with plane background or character images in clear environment. However, none of the methods are tested on character images captured by camera from natural scenes and video because these characters are complex in nature. In addition, thinning of different orientations of characters is not addressed well since arbitrary orientations make the problem more complex. Hence, in this work, we propose to explore ring radius transform (RRT) for obtaining thinned images for multi-oriented character images in video as well as natural scenes. This approach falls in the category of medial axis-based methods. As we are inspired by the work proposed in [17] for video character reconstruction using RRT, which is introduced to identify medial axis, we explore the same RRT for obtaining skeleton in this work. The existing RRT is good for horizontal and vertical gaps filling and the characters with small gaps but not arbitrary orientation characters. To fill the gaps in different directions, medial axes are obtained with different directions of RRT in [18], which is for scene character shape reconstruction and the directions are limited to few but not arbitrary directions. The improved version of the method is proposed for the reconstruction of video characters in [19] using iterative midpoint procedure. This method is good for small gaps but fails when the contour has multiple big gaps. Meanwhile, the above reconstruction methods are developed for video and scene characters from natural scene images but not for the purpose of thinning. Besides, the methods have not considered arbitrary orientations of characters, handwriting and multi-scripts. Therefore, the proposed work explores the same RRT to identify medial axis for arbitrary orientations in a novel way and obtain thinned images that preserve visual topology to achieve a good recognition rate. The main contributions of the proposed method are as follows. The novelty lies here is to extend the existing RRT concept for obtaining arbitrary skeletons without scarifying the shape of a character image. To achieve this, the proposed method combines the gradient direction and the direction given by principal component analysis (PCA) based on the neighbors for each pixel in a novel way to handle the problem of arbitrary orientations. The method proposes an iterative procedure to fill missing medial axis pixels based on neighbor information especially at corners, junctions and end points. Color information is also explored to eliminate false medial axis segments. We modify the method in [19] to fill the gaps between two medial axis segments. Finally, the idea for shape restoration is introduced to obtain the actual shape of a character from medial axis values. The main advantage of this step is to produce a complete shape of the character despite missing a few medial axis pixels. The above- mentioned steps are new compared to the existing methods [17–19]. Hence, the proposed method is different from the above reconstruction methods in terms of scope, objective, contribution and experimental results.

3 Proposed approach

This approach requires segmented characters from a video image as the input for thinning. We use our previous method [20] for segmenting characters from video text lines. This method explores Gradient Vector Flow (GVF) and the cost for cuts to identify the space between characters. The main advantage of this method is that it segments characters from text lines directly without segmenting words. In addition, the method works well for multi-oriented text lines and is robust to noise, low resolution, background variations, and font or font size variations. Therefore, we propose to use this method for character segmentation. For each segmented character, the proposed method uses Canny edge operator to get an edge image. This is because Canny edge detector is good for both low and high resolutions compared to the other edge detectors such as Sobel and Prewitt. Note that Canny edges are represented by white color against dark background. It is evident from [17] that Canny gives fine edges for individual characters in video of different variations compared to other edge operators like Sobel, Prewitt, etc. However, due to complex background of video, Canny gives lots of spurious edges. Therefore, we propose the steps to restore missing information to obtain the skeleton with the visual topology of the original shape of the character in subsequent sections.

We propose to use the same RRT (more details can be found in Sect. 3.1) in a different way to find the medial axis for a character in an arbitrary orientation. The basis for this novel idea is that the perpendicular direction to the contour pixels gives an actual direction for identifying the medial axis. Sometimes, the above procedure may disconnect at corner, junction and intersection points due to arbi-

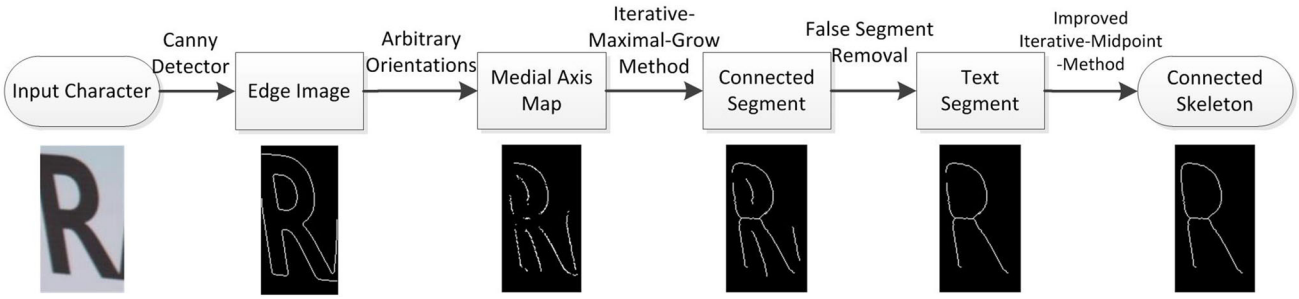


Fig. 2 Flow diagram of the proposed thinning method

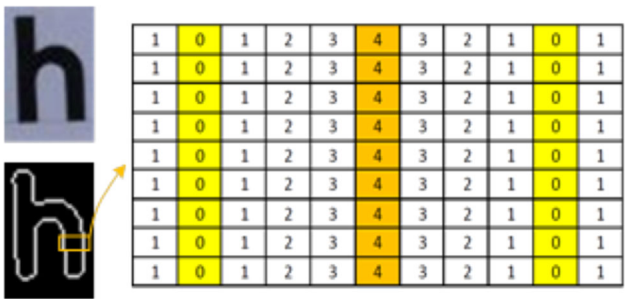


Fig. 3 Illustration of RRT for medial axis values: here "0" represents edge information and "4" represents medial axis values

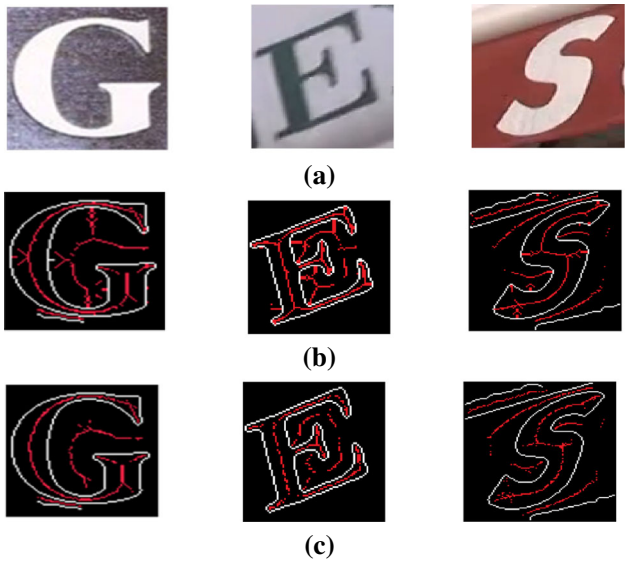


Fig. 4 Medial axes of the proposed and existing RRT. **a** Input video character image of different orientations, **b** medial axes of the existing RRT method [17], **c** medial axes of the proposed method

trary orientations. We propose an iterative-maximal-growing (IMG) method for connecting such gaps based on the direction of neighbor radius information to predict medial pixels. This may result in false medial axis segments due to background complexity. We remove false medial axis segments based on color and clustering combinations to identify stable medial axes segments. This idea works based on the fact that the actual medial axis has uniform distribution of colors compared to the medial axis created by backgrounds. Then we propose to use the modified iterative-midpoint-method (IMM) to fill the gap between two medial axis segments. Motivated by the work presented in [19] for character gap filling, we propose the same with modifications to fill the gap between medial axis segments. Finally, we restore the shape of the character using the radius values of medial axis pixels in a novel way. The shapes of the restored characters are sent to OCR for recognition. This will ensure the preservation of shape and visual topology. The steps of the proposed method for thinning can be seen in Fig. 2.

3.1 Ring radius transform for medial axis

For a given segmented character image, we produce its corresponding edge maps as the input of RRT. In general, RRT assigns a value to each pixel of an edge map based on its distance to the nearest edge pixel. The assigned radius value can be defined as follows:

$$rad\,(p) = \min_{edge(q)=1} dist(p, q)$$

where edge pixels and background pixels are assigned values 1 and 0, respectively, in an edge map *edge*, and $dist(p, q)$ represents the Euclidean distance between pixels p and q. Figure 3 shows one example of character 'h' with its corresponding Canny edge image and radius map. In Fig. 3, we can notice that the edges marked yellow color have radius value '0', and the radius value increases from zero to the highest radius value ('4') and then decreases to zero again between the inner side and the outer contour. These highest radius values marked with orange color denote the medial axis pixels, which can be used for obtaining the skeleton of the character. More details can be found from [17].

Since the existing RRT is limited to horizontal or vertical medial axis, it does not have the ability to find the medial axis of any oriented character as illustrated in Fig. 4, where the input video character images of different orientations are shown in Fig. 4a, while RRT produces more noisy medial

axis pixels especially at the corners and the end of segments as shown in Fig. 4b because the existing RRT is not invariant to rotations. On the other hand, the proposed method gives medial axis pixels without many noises as shown in Fig. 4c because the proposed modified RRT is able to find the direction of medial axes for any rotated character in a novel way.

3.2 Medial axis for multi-oriented characters

For each pixel in the Canny edge image, we perform ring radius Transform [17] to obtain a radius map, which contains radius values defined as the distances to the nearest edge pixel as shown in Fig. 5, where yellow color denotes edge pixels and orange color denotes medial axis pixels between two strokes. The medial axis pixel (MAP) is defined as the middle pixel of two strokes (inner and outer contours of an edge image). In order to find such medial axis for any orientation of a character image, the method proposes the following algorithm. The method determines a perpendicular direction to edge pixels based on its neighbor pixels as shown in Fig. 5, where the direction from an edge pixel p toward another edge pixel q on the opposite contour is the actual direction of the medial axis. The method finds radius values using RRT for the pixel between the two edge pixels p and q. Then along the same direction, the method finds the maximum distance (radius value), i.e., dp shown in Fig. 5 between the two edge pixels, which is defined as medial axis pixel (MAP).

Specifically, for each edge pixel p, we define a group of nearby connecting edge pixels in its neighboring area. Then we use PCA to calculate the direction d_{neigh} of this area, which is regarded as the orientation of this edge pixel. In other words, d_{neigh} gives the direction based on neighbor information. We choose PCA and neighbor pixel information to find the orientation rather than choosing the gradient direction as used in the past for text detection [21] to save the number of computations. In addition, we believe that Canny edge is more reliable than gradient information. After that, we calculate $d_p = -\frac{1}{d_{\text{neigh}}}$, which is roughly perpendicular to the orientation of the stroke in p. Then, we follow the ray $r = p + s \cdot d_p$ until the ray comes across another edge pixel q. Since it is hard to decide the right direction from the outer contour to the inner contour, we increase and decrease the value of s to reach q and q', respectively. The method finds the maximum radius in the area between $[p, q]$, $[p, q']$ pixels as MAP.

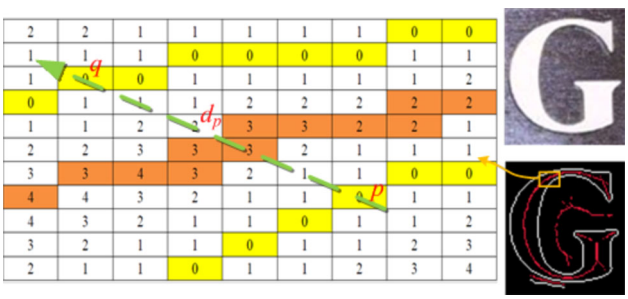


Fig. 5 Illustration for finding medial axis pixels in radius map: "p" and "q" are representing two edges pixels and "d_p" represents medial axis pixel which gives highest radius among all the pixels present between "p and "q"

The above procedure finds medial axes for the input character image as shown in Fig. 5, where medial axis pixels are found for character "G". It is found that the method finds medial axis pixels not only between strokes but also inside the character as in Fig. 5, where one can see red pixels inside the character. This results in noisy medial axis pixels as these medial axis pixels do not contribute for finding the skeleton of the character image. Therefore, to remove such noisy medial axis pixels, we propose the following criteria. From the radius map shown in Fig. 5, we can notice that the value of radius increases gradually from 0 (edge pixel). It reaches the maximum radius value (MAP), then decreases gradually and finally reaches 0 (edge pixel). This criterion helps to determine actual medial pixels.

The above criterion may give more than one medial axis pixels due to stroke width variations for some characters. To select the correct medial axis pixels, we perform histogram operation on the medial axis pixels found by criterion-1 and then we choose the radius which gives the highest peak as the real medial axis pixel. This is true because usually we expect a nearly constant stroke width throughout the character.

If the medial axis values satisfy the above two criteria, then the method considers them as actual medial axis values. The effect of these two criteria can be seen in Fig. 6, where the medial axis pixels inside character "G" has been successfully removed. Similarly, for other characters as shown in Fig. 6, we can see some of the noisy medial axis pixels are also removed compared with the results shown in Fig. 4. However, we can still see some noisy medial axis pixels in all the characters in Fig. 6, especially inside the characters and at convex areas.

3.3 Iterative-maximal-growing for contour completion

The local maximal values in the radius map give an estimation of character skeleton; however, in Fig. 6 we can notice disconnected medial axis segments. Such disconnections cannot be avoided due to the fact that they are probably caused by

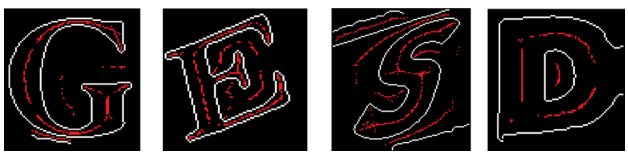


Fig. 6 Filtering noisy medial axis pixels with two criteria

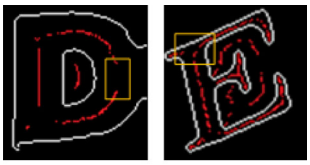


Fig. 7 Disconnected medial axis segments which are probably caused by **a** a stroke gap or **b** a stroke intersection

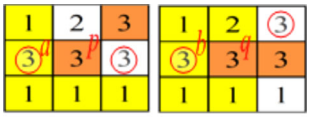


Fig. 8 The two situations of 8-connected MAP neighbors of a disconnection MAP: **a** the disconnection MAP p has a diagonal 8-connected neighbor with the same *orange color*, and **b** p has a 4-connected MAP neighbor. For the current disconnection MAPs p and q, we select a pixel from its *yellow* neighbors for skeleton growing to avoid the mergence with the existing diagonal 8-connected and 4-connected MAPs, respectively (color figure online)

stroke gaps or intersections inherently, which are shown in the yellow rectangles in Fig. 7. Thus, we propose an iteratively growth strategy to reconstruct a continuous skeleton for each character in this section. That is, we first search for all the disconnected MAPs as the seeds for growing, and then for each seed MAP, we select a proper neighboring pixel as the growth direction. This pixel will be considered as a new MAP for iteratively growing, which is repeated until an ending condition is met.

Specifically, we regard the MAPs generated in Sect. 3.1 as the initialized skeleton of a character, and identify each MAP p in them as a disconnection if the following condition is met:

$$|p_{\text{nei}}| = 0 \, or \, 1$$

where p_{nei} denotes the eight-connected neighbor MAPs of p and $|p_{\text{nei}}|$ gives the number of the eight-connected neighbors. That means, any MAP that has more than one eight-conneted MAP neighbors will be ignored during contour growing in our method. Thus, for any MAP, essentially there are altogether two situations for growing as, respectively, illustrated by Fig. 8a, b, in which the two disconnection MAPs p and q have a diagonal eight-connected neighbor MAP and a four-connected neighbor MAP with the same orange color, respectively. Note that for a MAP that has no eight-connected neighbor MAPs can be finally converted to one of the two situations by simply growing along with non-MAP pixels which have the maximal radius value.

For each identified disconnection MAP, we further search for all the non-MAP neighbor pixels that correspond to the local maximal radius value and select a proper pixel as the growth direction. This is because the growth of any disconnection MAP theoretically toward the maximal radius value direction according to the definition of our MAP. However, we still need select a proper direction since there may exist more than one pixels that correspond to the same local maximal radius value. For example, it can be seen in Fig. 8a that the two pixels of both the left and the right neighbors of p share the same maximal radius value "3"; however, the right pixel should not be selected to avoid the mergence with an existing MAP. For this purpose, we first collect all the pixels corresponding to the maximal radius value and consider them as candidates for growing. Then, for the current disconnection MAP, we select the only one pixel from the candidates on the condition that the pixel is not a 4-connected neighbor of the diagonal 8-connected MAP (situation 1 in Fig. 8a) or the 4-connected MAP (situation 2 in Fig. 8b). Thus, the mergence with the existing MAPs during skeleton growing can be avoided. As a result, in Fig. 8a, the pixel a will be finally selected for growing from disconnection p, and similarly the pixel b is selected for disconnection q in Fig. 8b. Note that sometimes the pixels corresponding to not only the maximal radius value but also the second maximal radius value can be both considered as candidates for more robust growth.

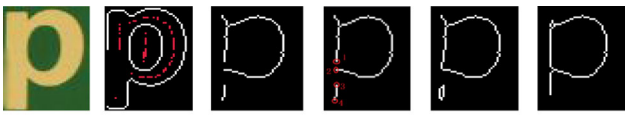
Fig. 9 Connecting medial axis segments using IMG

After the current MAP p is proceeded, the selected pixel will be considered as the current MAP for further growing. The growth will be stopped if any of the following two conditions is met: (1) None of a proper pixel can be successfully found during iteration, or (2) the growth becomes too close to any existing edge pixel. Our method works well for most disconnection cases caused by stroke gaps or stroke intersections. Figure 9 shows the results after iterative maximal growth from Fig. 6.

3.4 False segment removal

It is noted from the results shown in Fig. 9 that there are false medial axis segments. To eliminate such false medial axis segments, we propose a new method based on gray information of medial axis pixels. It is true that the gray values of actual medial axis pixels have uniform values compared to the gray values of other medial axis pixels created by background. This is because those values that represent edge pixels are usually constant, while the values that represent background will have lots of variations due to background variations. Therefore, we calculate the mean and the vari-

Fig. 10 Filtering false medial axis segments



(a) **(b)** **(c)** **(d)** **(e)** **(f)**

Fig. 11 Modified IMM: **a** input image, **b** medial axis results, **c** skeleton, **d** existing IMM, **e** result of existing IMM and **f** skeleton result of modified IMM

ance for the medial axis pixels of all the segments. We use k-means with $k = 2$ clustering algorithm to separate low variance values from high variance values. The cluster that gives a low mean is considered as actual medial axis pixels. The effect of this step can be seen in Fig. 10, where false medial axis segments are removed.

3.5 Improved iterative-midpoint-method for filling

Sometimes, though the previous method connects disconnected segments and gaps at corners, we may get gaps due to low resolution and complex background of video. Therefore, we propose to modify our previous method [19] which was developed for character reconstruction using iterative-midpoint-method (IMM). This IMM works well when the end pixel pair satisfies the mutual nearest neighbor criteria. As a result, this condition fails for the situations shown in Fig. 11, where (a) is the input image, (b) shows the results of medial axis given by the proposed method, (c) is the skeleton of the input image (medial axis pixels), (d) gives the situation for which the existing IMM fails: for the correct pair end points, namely (1, 2) and (3, 4), IMM fills the *gap* between (1, 2) end points but fails to connect the *small segment* with (3, 4) as the end points as shown in Fig. 11e. Figure 11f shows the proposed modifications to the existing IMM, which works well for such situation. In summary, the existing IMM does not connect small segments and only works well for small gaps on a large contour. To overcome the problem of the existing IMM, we propose modifications to the existing IMM with the following conditions to identify the correct pair of end pixels. Let p and q be the end points, the conditions are defined as

$$condition_1 = \begin{cases} 0 & \text{if} \quad con\,(p,q) \\ 1 & \text{if} \quad con\,(p,q) \quad \cap dis\,(p,q) < \alpha \cdot con\_len(p,q) \\ 1 & \text{if} \quad \neg con\,(p,q) \end{cases}$$

$$condition_2 = \begin{cases} 1 & \text{if} \quad |mean\,(preline\,(p,q)) - mean\,(stroke)| \leq adap(i) \\ 0 & \text{if} \quad |mean\,(preline\,(p,q)) - mean\,(stroke)| > adap(i) \\ adap\,(i) = 2 * mean\,(Grad\,(i)) \end{cases}$$

In condition$_1$, $con\,(p,q)$ is the function that connects two end points p and q, $dis\,(p,q)$ is the Euclidean distance between p and q, $con\_len(p,q)$ is a small incremental step from p to q, and α is a weight factor. The factor α is determined empirically (presented in Sect. 4). In condition$_2$, $preline\,(p,q)$ is the function that gets the initial line called preline using the existing IMM, $mean(x)$ is to find the mean of the gray values of the pixels of the stroke, $adap(i)$ is a self-adaptive parameter determined by the mean of Gradient values of image i. Basically, the above two conditions help us to connect two separated small segments based on gray values of strokes and the distance between the segments as the existing IMM fails to connect small segments. If the two end points satisfy the above two conditions, then the proposed method considers the two end points are a correct pair of end points. Thus the method uses the existing IMM to fill the gap for identifying two correct end points.

3.6 Shape restoration for recognition

The previous step gives the skeleton for the input image. In this work, skeleton is nothing but the medial axis pixels of the character image. For each medial axis pixel, we have its radius value, which is the distance between two strokes (inner and outer contours). Since it is the distance of the exact mid pixel of two contours, the method considers half of that distance as the distance between the inner contour to the medial axis pixel and the outer contour to the medial axis pixel. Then the method uses the same half distance to display white pixels for inner and outer contours restoration. This results in filled complete character as shown in Fig. 12, where the shapes of all the characters are successfully restored. The advantage of this restoration step is that it covers small gaps created by missing medial axis pixels. The result of this step goes to OCR, which is publicly available [22], to recognize the character.

In summary, we can conclude that though the proposed new ideas look like heuristics in appearance, they are objective for the proposed thinning framework because they are derived from actual facts and do not use any constant fixed thresholds. For instance, Sect. 3.2 presents the algorithm for the selection of actual medial axis values, which does not use

Fig. 12 Sample shape restored results of the proposed method

any threshold, instead it uses the gradual changes of radius values representing the highest peak in the histogram. Similarly, Sect. 3.3 proposes gradient direction information as well as neighbor radius values to restore missing pixels. Section 3.4 presents k-means clustering algorithm to eliminate false medial axis segments on the basis that false segments do not exhibit uniform color values, while actual medial axis segments have almost uniform color values. Additionally, Sect. 3.6 presents medial axis values for restoring the shapes of characters based on the radius values of medial axis pixels.

4 Experimental results

We divide the experimental section into three sub-parts, namely datasets and evaluation, experiment on skeleton detection to validate the skeletons given by the methods, and experiments on recognition to validate the shapes of the skeletons given by the methods.

4.1 Datasets and evaluation

In order to evaluate the proposed method, we consider 300 video characters from standard video database in ICDAR 2013 robust reading competition [23], 400 natural scene characters from ICDAR 2013 database, 450 natural scene characters from Street View Data (SVT) [24], 500 arbitrary-oriented characters from MSER-TD500 dataset [25] and 200 objects from standard object database (MPEG7) [5,26,27]. Apart from this, we also create our own data to test the multilingual ability of the proposed method, which includes 200 characters, namely 40 from each of the five scripts: Chinese, Korean, Arabic, Tamil and Japanese scripts. In addition, to test the effectiveness on handwritten characters, we also consider 100 characters from ICDAR 2013 Handwriting Segmentation dataset [28]. In total, 2,150 characters are considered for the purpose of experimentation, which include 500 characters from video, 1,350 characters from natural scene, 100 characters from handwritten character images and 200 from object database. Video data from ICDAR 2013 [23] suffer from low resolution and complex background due to the quality of video. The natural scene characters from ICDAR 2013 data usually have complex background and lots of variations in font, font style, etc. The natural scene characters from SVT data [24] are much more complex than ICDAR 2013 data because most of the images contain complex background with greenery, building, etc. The natural scene characters from MSRA-TD500 [25] are also complex as they involve arbitrarily oriented characters with complex background. The handwritten data from ICDAR 2013 handwriting contest data have different font styles [28]. The objects from MPEG7 database [5,26,27] have non-symmetrical structure compared to text data. Overall, we consider different categories of data to show that the proposed method is independent of data, script, contrast and application. This is the advantage of the proposed method compared to the exiting methods. Note that since our video data and object data do not provide ground truth, we calculate accuracy (values for the measures) manually.

We choose two recent state-of-the-art methods, namely the scale space filtering-based method [1] that uses various thinning approaches in an adaptive way to thin noisy sketch images, and the contour-based method [16] that proposes heuristics to use characteristics of text to obtain skeleton, for comparative studies. The methods are not tested on video, natural scene and objects data that suffer from low contrast and complex backgrounds. Therefore, the heuristics method may not work well for our data. In order to show the effectiveness of the proposed method, we choose two classical methods: (a) a fast parallel algorithm for thinning digital patterns by Zhang and Suen [4], which uses an iterative procedure to identify outside contour pixels based on neighboring connectivity, and (b) parallel thinning with two sub-iteration algorithms by Guo and Hall [3], which is the improved version of the method [4] and works well based on the same 8-connectivity. However, both of these two classical methods require binary scanned character images. Besides, these methods are sensitive to disconnections and distortions caused by low resolution and complex background. On top of this, the proposed method is compared with the RRT [17] and Tian et al. [18] since these two methods used similar concepts with different objectives. Further, we conduct experiments for the proposed method using Sobel and Prewitt edge detectors to show that the proposed method with Canny edge detector is better than Sobel and Prewitt detectors. This is mainly because Sobel and Prewitt are good for high contrast images, but not for low contrast images, while Canny is good for both high and low contrast images. Since the source codes are available for [1,3,4], we use the same source codes for experimentation and comparative studies. For other methods [16–18], we implement them for comparison.

We propose to use the measures as suggested in [1,16], which evaluate the performance of the method in terms of topology preservation, robustness and effectiveness. The measures are originally proposed for evaluating the skeletons of scanned binary character images [1,16]. In this work, we propose to use the combination of the same measures for evaluating the skeletons of video and scene character images. The reason is that a video/scene character image can either have plane background which likes the background in a scanned binary image, or complex background which contains various objects. In other words, video character image data are the mixture of both the two types of character images. Therefore, we propose the measures for validating both visual topology [1] and distortion effects [16] during thinning in this work. Specifically, for measuring the visual topology of the skele-

ton given by the proposed method, we define Measure-1 (M1) as in Eq. (1). This measure requires the area of the original image and the area of the resultant image. We calculate the area for the resultant image, which is formed by the maximal discs that fit to the original image along the skeleton as defined in [29]. This area is divided by the area of the Canny edge image of the input image as in Eq. (1). Similarly, one more Measure-2 for measuring visual topology is defined as in Eq. (2). This measure requires the number of the pixels in the resultant image (skeleton), say N_{ske}, and the base image (Canny of the input image), say N_b. This measure believes that the pixels of edges are nearly twice as those in the skeleton.

$$M_1 = \frac{Area[S']}{Area[S]} \tag{1}$$

$$M_2 = 1 - \left| \frac{1}{2} - \frac{N_{\text{ske}}}{N_b} \right| \tag{2}$$

In the above Eqs. (1) and (2), S' represents the area of the resultant image, while S is the Canny edge image of the input image. These two measures are used to test the topology of the skeleton given by the proposed method.

Usually, thinning or skeleton algorithms may not give good results at junction and intersection points due to orientation and heuristics constraints. Therefore, there is a necessity to measure the distortion at junction and intersection points. For this purpose, we define Measure-3 (M3), Measure-4 (M4) and Measure-5 (M5) as mentioned in Eqs. (3)–(5), respectively. For Eq. (3), if branches emanating from the junction point then we consider the junction point is distorted. In this way, we manually count both junction points (n_{jun}) from the Canny edge images and distorted junction points (n_{disj}) from the skeleton images to measure the robustness to distortions. For Eq. (4), the number of end points (n_{end}) is counted from the Canny edge image of the input image and the distortion end points from the skeleton image of the proposed method. We also use one more Measure-5 (M5) to measure the distortions due to spurious strokes as the fraction of the number of spurious strokes (n_{spurs}) at high curvature regions to the total number of strokes (n_{stroke}) as in Eq. (5). We count spurious strokes and the total number of strokes as defined for Measures-3 and Measure-4. In summary, M1 and M2 are used to measure the topology of the skeleton of the image, while M3 to M5 are used to measure the robustness to distortions. More details about the measures can be seen in [1] and [16], respectively. In summary, we use two types of measures: M_1–M_2 are used to measure visual topology, while M_3–M_5 are used to measure distortion effects. For an ideal character, M_1–M_2 should give high values, while M_3 to M_5 should give low values. Otherwise, the measures most probably denote the loss of a shape, which theoretically in turn leads to a poor accuracy of OCR since the current OCR engine [22] used in this work is very sensitive to shape variations or distortions.

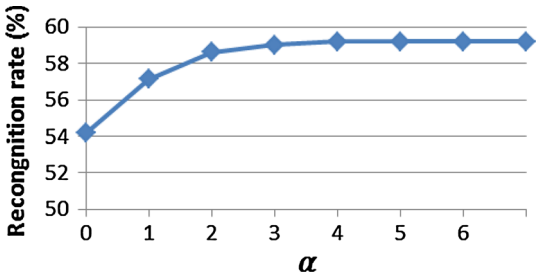


Fig. 13 Experimental study for determining value for α on 500 samples chosen randomly from all the datasets

$$M_3 = \frac{n_{\text{disj}}}{n_{\text{jun}}} \tag{3}$$

$$M_4 = \frac{n_{\text{disj}}}{n_{\text{end}}} \tag{4}$$

$$M_5 = \frac{n_{\text{spurs}}}{n_{\text{stoke}}} \tag{5}$$

Apart from the above measures, we also use recognition rate as a measure to evaluate the skeleton results given by the proposed method. If we achieve a good recognition rate for the obtained results, the method is said to be good to preserve the shape of the original character. For this purpose, we restore the shape from the skeleton (medial axis pixels) as we discussed in Sect. 3.6. Then we pass the restored results to OCR to calculate recognition rate. We calculate recognition rate before restoration (we send the originally input characters to OCR) and after restoration (we send the characters after restoring their shapes). This validates the effectiveness of the thinning or skeletonization.

For the threshold α used in Sect. 3.5, we randomly choose 500 images from our datasets to determine the optimal value. We plot a graph for recognition rate verses different α values as shown in Fig. 13. According to the experiments, the value for α is finally selected as 4, which is used for all the experiments in the work.

4.2 Experiments on skeleton detection

Sample qualitative results of the proposed and the existing methods including Sobel and Prewitt for thinning and recognition are, respectively, shown in Fig. 14a, b, where we can notice that the proposed method produces thinning results well for different situations like low contrast, complex background, blur and arbitrary orientations. On the other hand, the existing methods fail to obtain skeletons that preserve the shapes of character images. This is because all the existing methods except Shivakumara et al. [17] and Tian et al. [18] are developed for scanned images. However, Shivakumara et al. and Tian et al. have some inherent limitations and give poor accuracies.

The quantitative results of the proposed and the existing methods on different datasets are reported, respectively, from

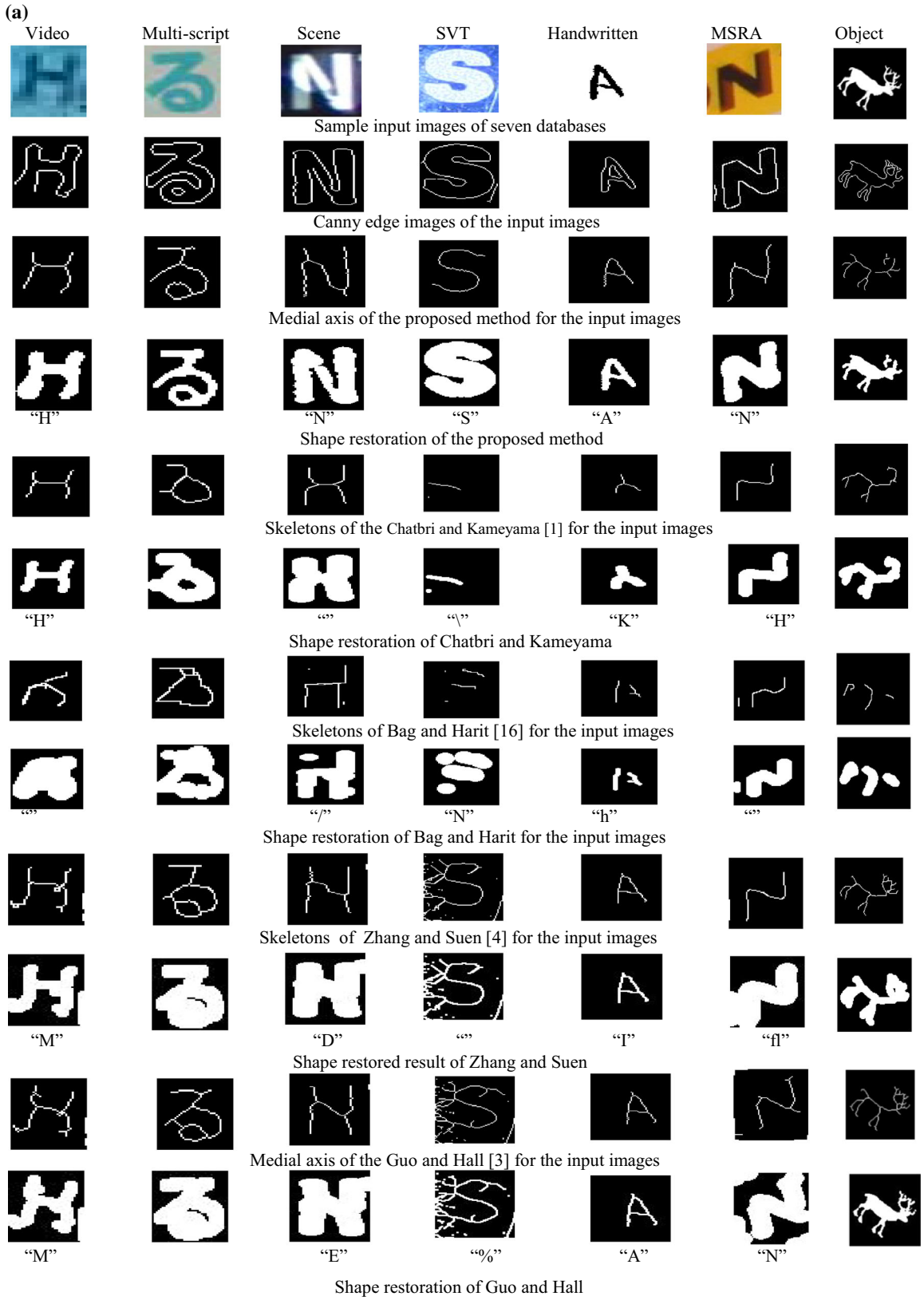


Fig. 14 **a** Sample skeleton and recognition results of the proposed and the existing methods. Note that recognition results are given under *double quotes*, and the *double quote without character* denotes no recognition result. **b** Sample skeletons and recognition results of the proposed and the existing methods. Note that the *double quote without character* denotes no recognition result (null)

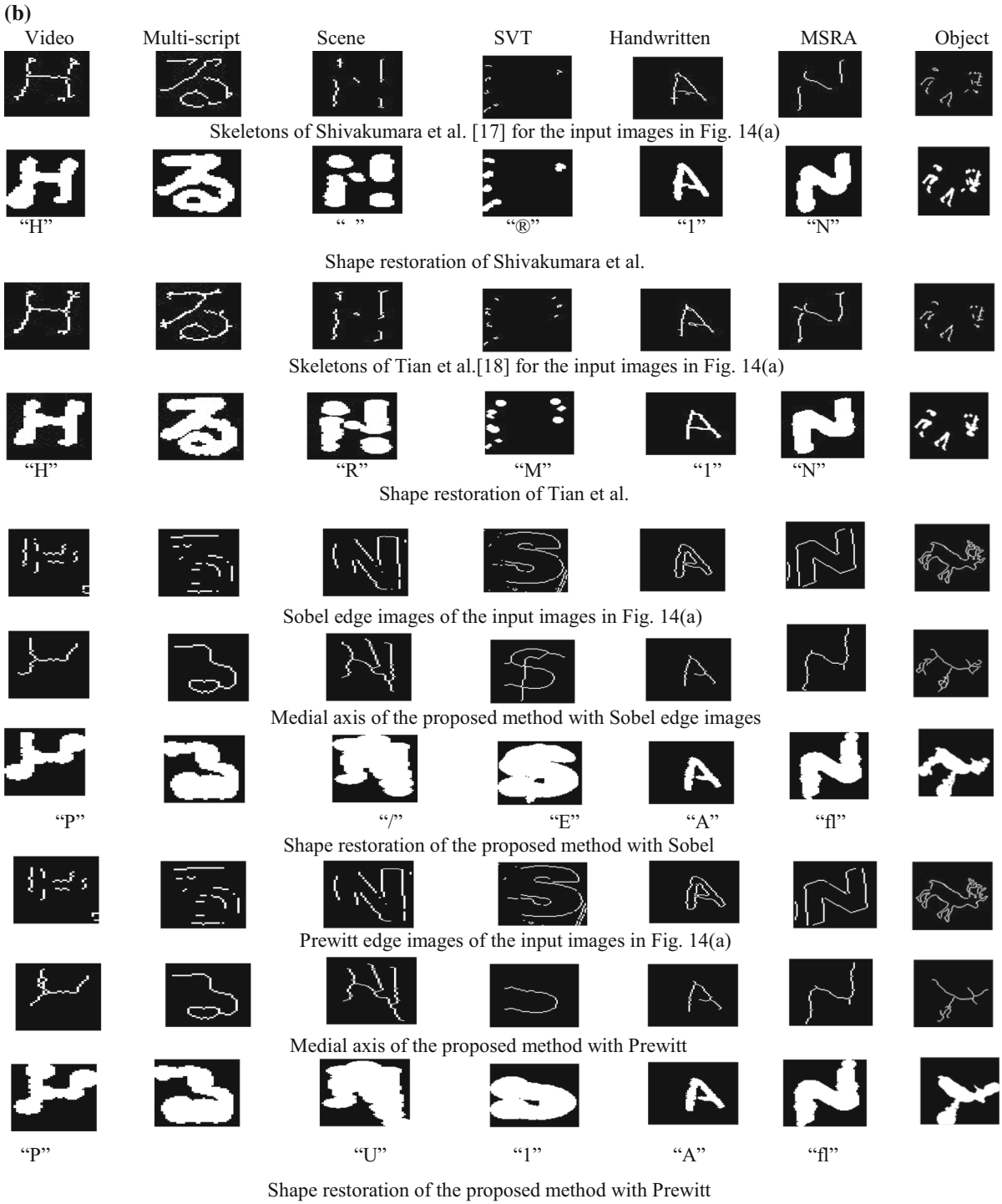


Fig. 14 continued

Tables 1, 2 and 3, where the results of the proposed method with different edge detectors are reported. It is observed from Tables 1, 2 and 3 that the proposed method with Canny edge detector is better than Sobel and Prewitt edge detectors. The main reason is that Sobel and Prewitt are not good for low resolution and complex background. It is observed from Tables 1, 2 and 3 that the two classical methods are better than other existing methods and the proposed method in terms of visual topology (M1 and M2), but the methods are worsen than the proposed method in terms of distortion as the methods give high values for M3–M5 due to the introduction of distortions during thinning. Shivakumara et al. [17] and

Table 1 Quality measures of the proposed and the existing methods on ICDAR-2013 video and multi-script video data

Methods	ICDAR 2013 video					Multi-script video				
	$\mathbf{M_1}$	$\mathbf{M_2}$	$\mathbf{M_3}$	$\mathbf{M_4}$	$\mathbf{M_5}$	$\mathbf{M_1}$	$\mathbf{M_2}$	$\mathbf{M_3}$	$\mathbf{M_4}$	$\mathbf{M_5}$
Chatbri and Kameyama [1]	0.54	0.81	0.13	**0.09**	**0.04**	0.19	0.52	**0.19**	**0.13**	**0.09**
Bag and Harit [16]	0.45	0.76	0.44	0.31	0.13	0.23	0.53	0.47	0.35	0.20
Zhang and Suen [4]	0.95	**0.90**	0.28	0.34	0.14	**0.80**	**0.92**	0.29	0.31	0.18
Guo and Hall [3]	**0.97**	**0.90**	0.50	0.47	0.30	0.76	0.91	0.41	0.44	0.30
Shivakumara et al. [17]	0.60	0.69	0.70	0.54	0.27	0.48	0.87	0.62	0.51	0.33
Tian et al. [18]	0.65	0.75	0.67	0.58	0.42	0.48	0.83	0.60	0.63	0.42
Proposed-Sobel	0.49	0.62	0.73	0.53	0.42	0.27	0.57	0.64	0.59	0.41
Proposed-Prewitt	0.48	0.65	0.72	0.54	0.40	0.30	0.50	0.63	0.56	0.41
Proposed-Canny	0.74	0.86	**0.13**	0.15	0.06	0.62	0.78	0.20	0.17	0.13

Bold values denote the best performance in the column

Table 2 Quality measures of the proposed and the existing methods on scene data of ICDAR 2013, SVT and MSRA

Methods	ICDAR 2013					SVT					MSRA-TD500				
	$\mathbf{M_1}$	$\mathbf{M_2}$	$\mathbf{M_3}$	$\mathbf{M_4}$	$\mathbf{M_5}$	$\mathbf{M_1}$	$\mathbf{M_2}$	$\mathbf{M_3}$	$\mathbf{M_4}$	$\mathbf{M_5}$	$\mathbf{M_1}$	$\mathbf{M_2}$	$\mathbf{M_3}$	$\mathbf{M_4}$	$\mathbf{M_5}$
Chatbri and Kameyama [1]	0.64	0.85	**0.11**	**0.03**	**0.02**	0.53	0.79	**0.09**	**0.03**	**0.01**	0.52	0.81	**0.24**	**0.12**	**0.05**
Bag and Harit [16]	0.62	0.82	0.19	0.20	0.11	0.47	0.75	0.33	0.34	010	0.35	0.53	0.25	0.15	0.08
Zhang and Suen [4]	0.80	**0.90**	0.21	0.17	0.09	**0.91**	**0.90**	0.24	0.36	0.07	**0.92**	**0.92**	0.24	0.10	0.03
Guo and Hall [3]	0.72	**0.90**	0.27	0.25	0.10	0.90	0.90	0.35	0.42	0.08	0.90	0.91	0.32	0.16	0.04
Shivakumara et al. [17]	0.59	0.70	0.19	0.23	0.16	0.61	0.69	0.6	0.63	0.22	0.71	0.70	0.30	0.41	0.22
Tian et al. [18]	0.64	0.77	0.44	0.44	0.35	0.69	0.76	0.63	0.70	0.34	0.76	0.76	0.33	0.46	0.29
Proposed-Sobel	0.70	0.80	0.20	0.41	0.30	0.47	−0.03	0.56	0.72	0.25	0.55	0.57	0.27	0.49	0.27
Proposed-Prewitt	0.69	0.77	0.30	0.44	0.30	0.45	−0.03	0.61	0.76	0.29	0.53	0.58	0.28	0.52	0.29
Proposed-Canny	**0.85**	0.81	0.18	0.05	0.03	0.70	0.88	0.14	0.10	0.03	0.72	0.86	0.25	0.16	0.09

Bold values denote the best performance in the column

Table 3 Quality measures of the proposed and the existing methods on handwritten and object data

Methods	Handwritten data					Object data				
	$\mathbf{M_1}$	$\mathbf{M_2}$	$\mathbf{M_3}$	$\mathbf{M_4}$	$\mathbf{M_5}$	$\mathbf{M_1}$	$\mathbf{M_2}$	$\mathbf{M_3}$	$\mathbf{M_4}$	$\mathbf{M_5}$
Chatbri and Kameyama [1]	0.30	0.80	**0.14**	0.08	**0.02**	0.70	0.82	**0.01**	0.01	0.05
Bag and Harit [16]	0.52	0.77	0.19	0.07	0.06	0.59	0.67	0.04	0.03	0.09
Zhang and Suen [4]	**0.90**	**0.95**	0.16	0.03	0.02	0.82	0.87	0.12	0.11	0.03
Guo and Hall [3]	0.81	0.92	0.21	0.03	0.02	**0.91**	**0.90**	0.03	0.02	0.03
Shivakumara et al. [17]	0.73	0.89	0.21	0.29	0.31	0.24	0.70	0.08	0.08	0.03
Tian et al. [18]	0.56	0.79	0.16	0.24	0.34	0.26	0.78	0.13	0.15	0.05
Proposed-Sobel	0.37	0.79	0.16	0.26	0.34	0.73	0.85	0.13	0.02	0.02
Proposed-Prewitt	0.34	0.78	0.21	0.29	0.33	0.73	0.84	0.14	0.05	0.06
Proposed-Canny	0.77	0.86	0.16	**0.05**	0.05	0.84	0.88	0.06	**0.01**	**0.02**

Bold values denote the best performance in the column

Tian et al. [18] are better than Chatbri and Kameyama [1] and Bag and Harit [16], but worsen than the proposed method in terms of visual topology and distortion free. Tables 1, 2 and 3 show that the method proposed by Chatbri and Kameyama is the best for the measures M3 to M5 (distortion free) compared to the other methods, including the proposed method for most of the experiments. However, it gives poor accuracies for M1–M2 (visual topology). As a result, there are losses for character shapes as shown in Fig. 14a, where we can see that restored results do not well preserve the original

Table 4 Recognition rates of the proposed and the existing methods on ICDAR 2013 video and handwritten data (BR: before restoration, AR: after restoration and IMT: Improvements)

Methods	ICDAR 2013 video			Handwritten data		
	BR (%)	AR (%)	IMT (%)	BR (%)	AR (%)	IMT (%)
Chatbri and Kameyama [1]	37.13	40.93	3.80	18.68	19.78	1.1
Bag and Harit [16]		16.03	−21.10		10.15	−8.53
Zhang and Suen [4]		25.32	−11.81		16.48	−2.2
Guo and Hall [3]		22.78	−14.35		24.18	5.5
Shivakumara et al. [17]		32.30	−5.17		17.37	−1.69
Tian et al. [18]		38.4	1.27		20.88	2.2
Proposed-Sobel		24.47	−12.66		12.09	−6.59
Proposed-Prewitt		24.47	−12.66		9.89	−8.79
Proposed-Canny		54.01	**16.88**		28.77	**10.09**

Bold values denote the best performance in the column

Table 5 Recognition rate of the proposed and the existing methods on ICDAR 2013 scene, SVT and MSRA data (BR: before restoration, AR: after restoration and IMT: improvements)

Methods	ICDAR 2013 scene			SVT			MSRA-TD500		
	BR (%)	AR (%)	IMT (%)	BR (%)	AR (%)	IMT (%)	BR (%)	AR (%)	IMT (%)
Chatbri and Kameyama [1]	51.50	59.00	7.50	40.74	41.90	1.16	39.84	39.25	−0.59
Bag and Harit [16]		36.25	−15.25		26.62	−14.12		19.92	−19.92
Zhang and Suen [4]		49.00	−2.5		33.80	−6.94		29.59	−10.25
Guo and Hall [3]		48.50	−3		29.86	−10.88		28.60	−11.24
Shivakumara et al. [17]		49.75	−2.25		42.86	2.12		41.31	1.47
Tian et al. [18]		53.25	1.75		52.31	11.57		49.11	9.27
Proposed-Sobel		55.25	3.75		14.58	−26.16		33.53	−6.31
Proposed-Prewitt		53.25	1.75		16.43	−24.31		32.74	−7.10
Proposed-Canny		67.75	**16.25**		71.53	**30.79**		50.30	**10.46**

Bold values denote the best performance in the column

shapes. This further leads to a poor recognition rate since the OCR engine is very sensitive to shape losses as introduced. On the other hand, the proposed method is better than all the existing methods excluding the two classical methods in terms of visual topology preservation and distortion free because of the advantages of RRT and the shape restoration method that uses medial axis pixel values. Thus, we get good recognition rates.

When we compare the results reported of ICDAR 2013 video data and multi-script video data in Table 1, the accuracy for multi-script data is lower than video data. This is because of the complex nature of the scripts especially like Chinese, Japanese and Koran, where high cursiveness can be expected. This may cause to get a poor accuracy.

4.3 Experiments on recognition

To validate whether the skeletons or the thinning given by the methods preserve the shapes of characters, we pass both the input and the restored characters to available OCR [22] to calculate the recognition rates before and after restoration. The sample recognition results for thinning are shown in Fig. 14a, b, where we can observe that the skeletons given by the proposed method are recognized correctly, while the skeletons given by the existing methods are not recognized correctly. This shows that the proposed method is good at preserving the shapes of the characters through medial axis formation. The quantitative results of the proposed and the existing methods on different datasets are reported in Tables 4 and 5. Note that there are no recognition results for multi-script and object datasets because of the non-availability of OCR publicly. Tables 4 and 5 show that the proposed method gives better recognition rates after restoration (AR) compared to before restoration (BR). Therefore high Improvements (IMT) are reported in the Tables by the proposed. Overall, the proposed method is better than all the existing methods in terms of recognition rate including Shivakumara et al. and Tian et al. This concludes that the skeletons given by the proposed

method preserve the shapes of the characters. Though the existing methods give good skeletons but fail to give good recognition rates because of noises and the distortion produced during thinning.

Overall, it is observed from the experimental results that the existing methods give reasonably good results, which are close to the results of the proposed method for obtaining skeletons, while for recognition results the existing methods drop down drastically compared to the proposed method. It is noted that Guo and Hall [3] and Zhang and Suen [4] are the best at M_1–M_2 compared to all other methods including the proposed methods in terms of the quality measures. However, when we look at the recognition results, the accuracy drops to the value which is lower than the proposed method. The reason is that the methods [3,4] are good at M_1–M_2 but not good at M_3–M_5 compared to the proposed method. That is, the two methods introduce distortion while obtaining skeleton for video and scene character images due to the variations in background. As a result, the two methods [3,4] lose the shape of a character image during restoring the shape from the skeleton. Hence, poor recognition rates for the methods [3,4] are achieved compared to the proposed method. This is because the proposed method has the advantage of RRT and the shape restoration method which preserves the shape of a character well. In addition, all the existing methods except Shivakumara et al. [17] and Tian et al. [18] used here for comparative studies are developed to obtain the skeletons of scanned character images or synthetic images. Since Shivakumara et al. and Tian et al. are limited to a few directions for finding medial axis, the methods give poor accuracies for arbitrary orientated characters. As a result, the methods lose the shapes of characters and introduce distortion for the character images from video and natural scene images. Therefore, the recognition accuracies for the skeletons given by these two existing methods are poor. In particular, [16] is inconsistent for all the experiments because this method is developed for synthetic and printed character images, where plane background and the clear shapes of characters exist. In addition, another reason to get poor recognition rates is that the current OCR limitation, such as the Tessearct OCR accepts high- resolution images with homogeneous background and little tilted. On the other hand, since the proposed method is developed for handling video images and naturals scene images of any orientation, it preserves the topology and the structure of such characters, which in turn helps in achieving a good recognition rate. In summary, the proposed method outperforms the existing methods for all the types of data in terms of distortion free and recognition rate.

It is also observed from Tables 4 and 5 that the OCR gives the lowest recognition rate before restoration for handwritten data compaerd to other data because the OCR is not rboust to the variatios in handwritting styles. Similarly, the recognition rate increases for other data in the order of video, MSRA, SVT, ICDAR 2013 scene data. The reasons behind this are video consists of low relution with complex background, MSRA consists of arbitrary orientations with complex background, SVT data consist of small fonts with complex backgrond, and ICDAR 2013 scene data have horiozntal direction with less complex background compared to MSRA and SVT. As a result, the different complexities of the data reflect in achieiving different recogniton rates by the existing methods. However, the proposed method gives almost consistent recogniton rates compared to the existing methods except for the handwrittn data. This is due to OCR limitations. Thus, we can conclude that the proposed method works well for different data with different complexities.

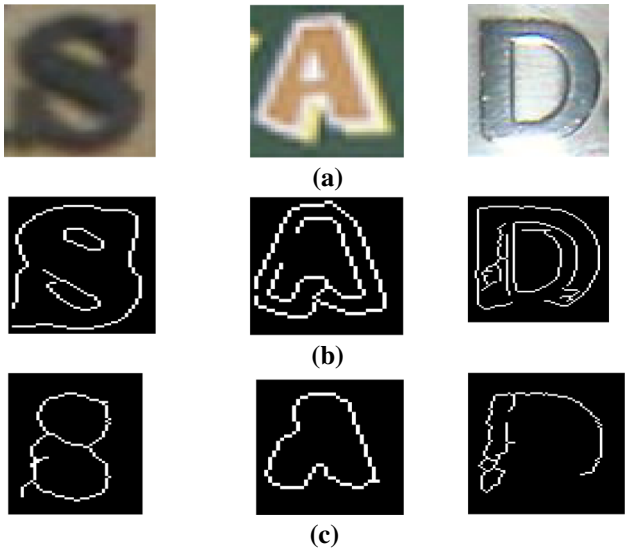


Fig. 15 Failure cases of the proposed method. **a** Input images, **b** Canny edge images, **c** skeleton obtained by the proposed method

4.4 Failure cases

Since the proposed method uses Canny edge image of the segmented character image for producing medial axis, sometimes, Canny edge detector gives poor results for character images as shown in Fig. 15a, where one can notice that the images are too blur, complex background and heavily affected illumination. In these situations, Canny edge detector gives poor edges as shown in Fig. 15b. Therefore, the proposed method fails to generate proper medial axis or skeletons as shown in Fig. 15c, where character images lose shapes. Thus, there is a scope for improvement in future.

5 Conclusion and future work

In this paper, we have proposed a new method based on ring radius Transform (RRT) for thinning. The RRT is pro-

posed to identify the medial axis pixels for arbitrary orientations of character images in a novel way. A new iterative-maximal-growing method is proposed to connect medial axis at junction, intersection and end points based on selecting maximal radius values. We propose to modify the existing iterative-midpoint-method (IMM) to fill the gaps at different segments of medial axes without losing shape. Experimental results on video, natural scene, arbitrary-oriented, multi-script, handwritten and object data show that the proposed method is superior to the existing methods in terms visual topology and shape preservation. However, the two classical thinning methods are better than the proposed method in terms of visual topology but worst in terms of distortion. Furthermore, the experimental results reveal that the proposed method is independent of data, type, script and object. Our future works would be the extension of this idea for object retrieval and recognition on large databases with complex shapes.

Acknowledgments The work described in this paper was supported by the Natural Science Foundation of China under Grant Nos. 61272218 and 61321491, and the Program for Chinese New Century Excellent Talents under NCET-11-0232. This research is also supported in part under Grant No. UM.TNC2/IPPP/UPGP/261/15 (BKP010-2013). We thank the anonymous reviewers for their constructive comments, which helped to improve the paper.

References

1. Chatbri, H., Kameyama, K.: Using scale space filtering to make thinning algorithm robust against noise sketch images. Pattern Recognit. Lett. **42**, 1–10 (2014)
2. Su, Z., Cao, Z., Wang, Y.: Stroke extraction based ambiguous zone detection: a preprocessing step to recover dynamic information from handwritten Chinese characters. In: IJDAR, pp. 109–121 (2009)
3. Guo, Z., Hall, R.W.: Parallel thinning with two-subiteration algorithms. Commun. ACM **32**(3), 359–373 (1989)
4. Zhang, T.Y., Suen, C.Y.: A fast parallel algorithm for thinning digital patterns. Commun. ACM **27**(3), 236–239 (1984)
5. Ward, A.D., Hamarneh, G.: The groupwise medial axis transform for fuzzy skeletonization and pruning. IEEE Trans. Pattern Anal. Mach. Intell. **32**(6), 1084–1096 (2010)
6. Alginahi, Y.M.: A survey on Arabic character segmentation. In: IJDAR, pp. 105–126 (2013)
7. Lam, L., Lee, S.-W., Suen, C.Y.: Thinning methodologies—a comprehensive survey. IEEE Trans. Pattern Anal. Mach. Intell. **14**(9), 869–885 (1992)
8. Sharma, N., Pal, U., Blumenstein, M.: Recent advances in video based document processing: a review. In: Proceedings DAS, pp. 63–68 (2012)
9. Zang, J., Kasturi, R.: Extraction of text objects in video documents: recent progress. In: Proceedings DAS, pp. 5–17 (2008)
10. Shivakumara, P., Phan, T.Q., Tan, C.L.: A Laplacian approach to multi-oriented text detection in video. IEEE Trans. Pattern Anal. Mach. Intell. **33**(2), 412–419 (2011)
11. Zhao, D., Shivakumara, P., Lu, S., Tan, C.L.: New spatial-gradient-features for video script identification. In: Proceedings DAS, pp. 38–42 (2012)
12. Phan, T.Q., Shivakumara, P., Ding, Z., Lu, S., Tan, C.L.: Video script identification based on text lines. In: Proceedings ICDAR, pp. 1240–1244 (2011)
13. Hoffman, M.E., Wong, E.K.: Scale-space approach to image thinning using the most prominent ridge line in the image pyramid data structure. In: Proceedings SPIE, pp. 242–252 (1998)
14. Cai, J.: Robust filtering-based thinning algorithm for pattern recognition. Comput. J. **55**(7), 887–896 (2012)
15. Chen,Y.-S., Yu, Y.-T.: Thinning approach for noisy digital patterns. Pattern Recognit. **29**(11), 1847–1862 (1996)
16. Bag, S., Harit, G.: An improved contour-based thinning method for character images. Pattern Recognit. Lett. **32**(14), 1836–1842 (2011)
17. Shivakumara, P., Phan, T.Q., Bhowmick, S., Tan, C.L., Pal, U.: A novel ring radius transform for video character reconstruction. Pattern Recognit. **46**(1), 131–140 (2013)
18. Tian, S., Shivakumara, P., Phan, T.Q., Tan, C.L.: Scene character reconstruction through medial axis. In: Proceedings ICDAR, pp. 1360–1364 (2013)
19. Shivakumara, P., Hong, D.B., Zhao, D., Tan, C.L., Pal, U.: A new iterative-midpoint-method for video character gap filling. In: Proceedings ICPR, pp. 673–676 (2012)
20. Phan, T.Q., Shivakumara, P., Lu, S., Tan, C.L.: A gradient vector flow-based method for video character segmentation. In: Proceedings ICDAR, pp. 1024–1028 (2011)
21. Epshtein, B., Ofek, E., Wexler, Y.: Detecting text in natural scenes with stroke width transform. In: Proceedings CVPR, pp. 2963–2970 (2010)
22. Tesseract. http://code.google.com/p/tesseract-ocr/
23. Karatzas, D., Shafait, F., Uchida, S., Iwamura, M., Boorda, L.G.I., Mestre, S.R., Mas, J., Mota, D.F., Almazan, J.A., De las Heras, L.P.: ICDAR 2013 robust reading competition. In: Proceedings ICDAR, pp. 1115–1124 (2013)
24. Phan, T.Q., Shivakumara, P., Tian, S., Tan, C.L.: Recognizing text with perspective distortion in natural scenes. In: Proceedings ICCV, pp. 569–576 (2013)
25. Yao, C., Bai, Z., Liu, W., Ma, Y., Tu, Z.: Detecting texts of arbitrary orientations in natural scene imags. In: Proceedings CVPR, pp. 1083–1090 (2012)
26. Latecki, L.J., Lakamper, R., Echardt, U.: Shape description for non-rigid shapes with a single closed conrour. In: Proceedings CVPR, pp. 424–429 (2000)
27. Jalba, A., Wilkinson, M.H.F., Roerdink, J.B.T.M.: Shape representation and recognition through morphological curvature scale spaces. IEEE Trans. Image Process. **15**(2), 331–341 (2006)
28. Stamatopoulos, N., Gatos, B., Louloudis, G., Pal, U., Alaei, A.: ICDAR2013 Handwriting Segmentation Contest. In: Proceedings ICDAR, pp. 1402–1406 (2013)
29. Jang, B.-K., Chin, R.T.: One-pass parallel thinning: analysis, properties, and quantitative evaluation. IEEE Trans. Pattern Anal. Mach. Intell. **14**(11), 1129–1140 (1992)