Int J Adv Manuf Technol
DOI 10.1007/s00170-009-2250-1

ORIGINAL ARTICLE

Part-based model for visual detection and localization of gas tungsten arc weld pool

Fanhuai Shi · Xixia Huang · Ye Duan · Shanben Chen

Received: 9 May 2009 / Accepted: 2 August 2009

Abstract Observing the weld pool and measuring its geometrical parameters are important issues for automated and robotic welding, wherein the visual detection and localization of weld pool are critical steps. Previous methods of visual measurement of weld pool usually assume that the weld pool exists in a predefined area, and its contour should be a specific geometric shape and size. Furthermore, previous methods were only suited for the pool images with complete boundary information and with small disturbing/noise edge. When part of the pool boundary is seriously spoiled (for example, by reflection) or confused by pileup area, it is very difficult if not impossible to conduct the geometric measurement of weld pool. In this paper, we propose a robust visual detection and localization method for the pool of gas tungsten arc weld based on part-based modeling and recognition of objects. It provides an elegant framework for representing the outline of a weld pool and is especially efficient for weld pool detection and localization in cluttered scenes, when it is partially occluded or when similar-looking pileup area can act as additional distracters. Experiments on real images verified the proposed method.

Keywords Part-based model · Object detection · Visual localization · Arc welding · Weld pool

F. Shi (✉) · S. Chen
Welding Engineering Institute, Shanghai Jiao Tong University,
800 Dongchuan Road,
Shanghai 200240, China
e-mail: fhshi@sjtu.edu.cn

X. Huang
Marine Technology and Control Engineering Key Laboratory,
Shanghai Maritime University,
1550 Pudong Dadao,
Shanghai 200135, China

Y. Duan
Computer Science Department, University of Missouri-Columbia,
Columbia, MO 65201, USA

1 Introduction

In recent decades, more and more vision-based technologies are studied and applied for automated welding. More specifically, vision-based sensors acquire images from the weld pool, followed by image processing to analyze and extract related geometric parameters of the weld pool, which can then be used as feedback to a control system to automatically adjust the welding parameters accordingly [1–6].

Early works on vision sensor-based direct view of weld pool use a specially designed camera whose high-speed shutter is synchronized with a short-duration pulsed laser [2]. This camera allows the arc light to be eliminated from the image. However, since the camera is specifically designed, it has a much higher cost. In order to eliminate the influence of arc light on images acquired by standard off-the-shelves low-cost camera, researchers fixed a spectrum filter and a dimmer glass in front of the charge-coupled device (CCD) camera in pulsed gas tungsten arc weld (GTAW) [3, 7] or continuous GTAW [8]; thus, the particular wavelength does not pass through the spectrum filter. As a result, the two-dimensional weld pool region is clearly imaged.

In order to conduct edge localization of a weld pool, Wang et al. [7] assumed that the location and the size of the pulsed GTAW pool are nearly constant. They adopted the backpropagation neural network to remove the noise edge

Published online: 17 August 2009

pixel and took the remnant edge pixels as the pool edge, in the case of continuous GTAW, since the arc light is too strong to be eliminated by the filter. Shen et al. [8, 9] obtained the parameters of the weld pool indirectly by measuring the arc column area and assuming that there is only one major curve edge in the predefined fixed-size image window. In [10], Song and Zhang obtained the pool boundary information by their special configuration and constructed a piecewise boundary model for the weld pool.

Note that existing methods often assume that the weld pool exists in a predefined area and its contour to be a specific geometric shape and size. Furthermore, existing methods are only suited for situations such that the pool image has complete boundary information and has very small disturbing edge (i.e., with little reflection on the weld pool boundary area). When part of the pool boundary is seriously spoiled (for example, with serious reflection on the weld pool boundary) or confused by pileup area (e.g., Fig. 1), it would be very difficult if not impossible for existing methods to recover the complete edge information of the weld pool, thus failing to measure the geometric parameters of weld pool. Hence, it is imperative to develop more robust visual detection and localization methods to detect and localize the weld pool even when its boundary was seriously corrupted by the reflection. This paper aims to serve this need based on part-based modeling and recognition of objects.

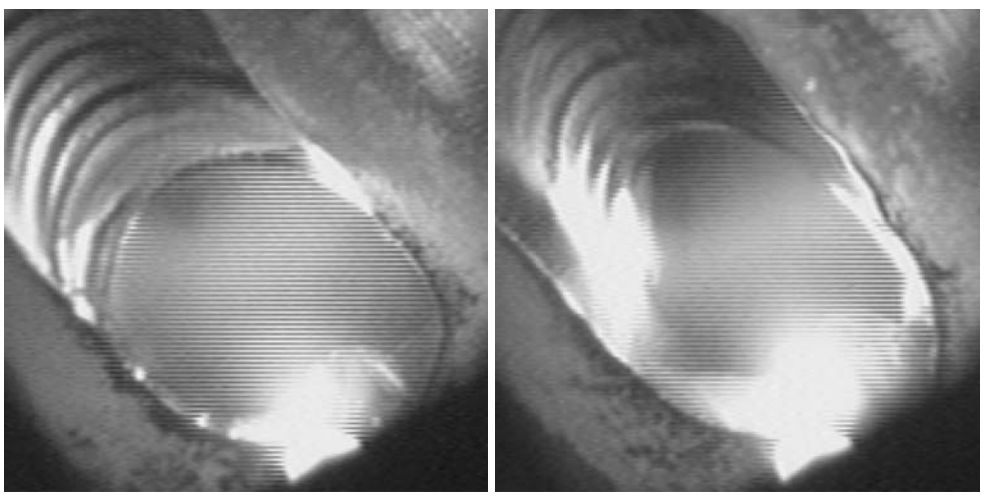
The remainder of this paper is organized as follows. Section 2 will introduce the principle of part-based models for object detection and localization. Section 3 describes the robust detection and localization of GTAW weld pool by part-based model. In Section 4, we will conduct some experiments to demonstrate the performance of the proposed method.

2 Part-based model for object detection and localization

This notion of part-based model originally appeared as the pictorial structure models introduced by Fischler and Elschlager [11] in 1973 and was recently adopted by many researchers including Felzenszwalb and Huttenlocher [12] and Leibe et al. [13]. Part-based model provides an elegant framework for representing object categories and is especially efficient for object detection and localization in cluttered real-world scenes where objects are often partially occluded and similar-looking background structures can act as additional distractions. To make the paper self-contained, we will briefly overview the main principles of part-based model for object detection and localization in this section.

In part-based model, an object is modeled by a collection of parts arranged in a deformable configuration. Each part encodes local visual properties of the object, and the deformable configuration is characterized by spring-like connections between certain pairs of parts. The best match of such a model to an image is found by minimizing an energy function that measures both a match cost for each part and a deformation cost for each pair of connected parts [12]. For example, in Fig. 2, the feature parts of a face include hairs, eyes, nose, mouth, etc., and the spring-like connections allow for variation in the relative locations of these features.

Depending on the different connectivity structures used to represent the components, part-based model can be further divided into several categories including constellation model, star-shaped model, k-fan model, tree model, bag of features, etc. [14]. For illustration purpose, in this section, we will use the tree-structured models of [12] as an example to describe the model representation, parameters learning, and model matching of part-based model.



(a) (b)

Fig. 1 Sample images of weld pool captured in the same configuration. **a** Image with little reflection. **b** Image with serious reflection on the weld pool boundary

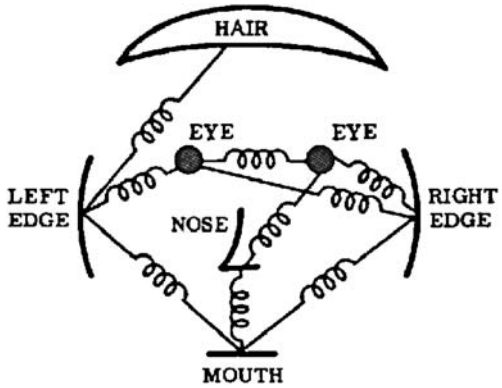


Fig. 2 The feature parts of a face [11]

2.1 Model representation

A particular modeling scheme must define the pose space for the object parts, the form of the appearance model for each part, and the type of connections between parts. In this subsection, we will describe models that represent objects by the appearance of local image patches and spatial relationships between those patches.

In this class of models, the location of a part is specified by its (x, y) position in the image, so we have a two-dimensional pose space for each part. To describe the appearance of each individual part, we can use the iconic representation introduced by Rao and Ballard [15]. More specifically, a high-dimensional vector is defined at the center position of an image patch to collect all the responses from a set of image filters at that point. This vector is normalized and called the iconic index at that position.

The appearance of a part is then modeled by a distribution over iconic indices. Particularly, we model the distribution of iconic indices at the location of a part as a Gaussian distribution with diagonal covariance matrix. Under the Gaussian model, the appearance parameters μ_i for each part include a mean vector μ_i and a covariance matrix Σ_i. We have

$$p(I|l_i,u_i) \propto N(\alpha(l_i),\mu_i,\Sigma_i), \quad u_i=(\mu_i,\Sigma_i)$$

where $\alpha(l_i)$ is the iconic index at location l_i in the image. We can easily use the mean and covariance of the iconic indices corresponding to the positive examples of a particular part to estimate the maximum likelihood (ML) parameters of this distribution in the following section.

The spatial configuration of the parts is modeled by a collection of springs connecting pairs of parts. Each connection (v_i, v_j) is characterized by the ideal relative location of the two connected parts s_{ij}, and a full covariance matrix Σ_{ij} which in some sense corresponds to the stiffness of the spring connecting the two parts. Let us define connection parameters as $c_{ij}=(s_{ij}, \Sigma_{ij})$; we can then model the distribution of the relative location of part v_i with respect to the location of part v_j as a Gaussian distribution with mean s_{ij} and covariance Σ_{ij}:

$$p(l_i,l_j|c_{ij}) = N(l_i-l_j,s_{ij},\Sigma_{ij}) \tag{1}$$

Ideally, the location of part v_i would be the location of part v_j shifted by s_{ij}. However, since the models are deformable, the location of v_i can vary by paying a cost that depends on the covariance matrix which corresponds to stretching the spring. Because we have a full covariance matrix, stretching in different directions will have different costs.

In practice, the joint distribution of l_i and l_j usually needs to be written into another Gaussian distribution with zero mean and diagonal covariance. Assume the singular value decomposition of the covariance matrix $\Sigma_{ij}=U_{ij}D_{ij}U_{ij}^T$ and define the following transformations

$$T_{ij}(l_i)=U_{ij}^T(l_i-s_{ij}) \quad \text{and} \quad T_{ji}(l_j)=U_{ij}^T(l_j)$$

Then, we can write Eq. 1 in the new form of

$$p(l_i,l_j|c_{ij}) = N(T_{ij}(l_i)-T_{ji}(l_j),0,D_{ij}) \tag{2}$$

2.2 Learning model parameters

Given a set of example images $\{I^1, ..., I^m\}$ and the corresponding object configurations $\{L^1, ..., L^m\}$ for each image, we want to estimate the model parameters, $\theta=(u,E,c)$, where $u=\{u_1, ..., u_n\}$ are the appearance parameters for each part; E is the set of connections between parts, and $c=\{c_{ij}|(v_i,v_j)\in E\}$ are the connection parameters. The ML estimate of θ is, by definition, the value θ^* that maximizes

$$p(I^1,\ldots,I^m,L^1,\ldots,L^m|\theta)=\prod_{k=1}^{m}p(I^k,L^k|\theta)$$

where the right-hand side is obtained by assuming that each example was generated independently. Since

$$p(I,L|\theta)=p(I|L,\theta)p(L|\theta),$$

the ML estimate is

$$\theta^*=\arg\max_{\theta}\prod_{k=1}^{m}p(I^k|L^k,\theta)\prod_{k=1}^{m}p(L^k|\theta). \tag{3}$$

The first term in Eq. 3 depends only on the appearance of the parts, while the second term depends only on the set of connections and connection parameters.

2.3 Matching with model

In this subsection, we will briefly introduce an efficient algorithm for matching tree-structured model to images

with connections of the form in Eqs. 2 and 4. Given an image, let $m_i(l_i)$ be a cost function that measures the degree of mismatch when part v_i is placed at location l_i in the image, and assume that $d_{ij}(l_i, l_j)$ is the Mahalanobis distance between transformed locations,

$$d_{ij}(l_i, l_j) = \left(T_{ij}(l_i) - T_{ji}(l_j)\right)^T M_{ij}^{-1}\left(T_{ij}(l_i) - T_{ji}(l_j)\right) \quad (4)$$

The algorithm solves the following energy minimization problem

$$L^* = \arg\min_L \left(\sum_{i=1}^{n} m_i(l_i) + \sum_{(v_i,v_j)\in E} d_{ij}(l_i, l_j) \right) \quad (5)$$

which is a configuration minimizing the sum of the match costs m_i for each part and the deformation costs d_{ij} for the connected pairs of parts. It is equivalent to finding the maximum a posteriori estimate of the object location in the statistical framework given an observed image.

Solving the minimization problem in Eq. 5 for arbitrary graphs and arbitrary functions m_i, d_{ij} is NP-hard. However, by restricting the graphs to trees and with the restricted form for d_{ij} shown in Eq. 4, the problem can be solved more efficiently [12]. The overall running time of this algorithm is $O(N^2)$.

3 Detection and localization of weld pool by part-based model

3.1 Model design and training

In this paper, we developed a part-based model for visual detection and localization of GTAW pool. More specifically, we divide the appearance model of a weld pool image into five parts $v_i(i = 1 \ldots 5)$, top, top-left, top-right, bottom-left, and bottom-right (Fig. 3a), and construct a five-part model based on the following rationales:

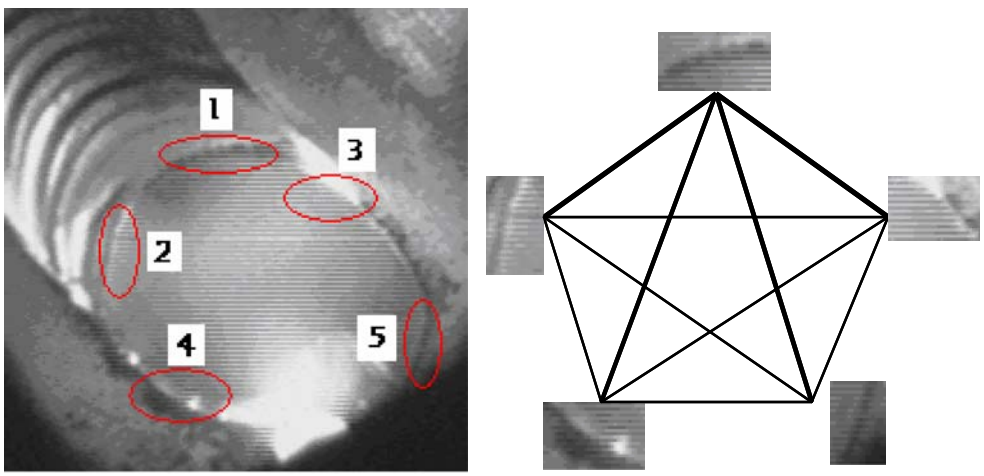
1. The central area of a GTAW pool is usually uniform, and most of the distinctive appearance features are distributed around its boundary contour.
2. The shape of the boundary contour of the weld pool is usually convex.
3. The position of part 1 is usually one of the most important reference points for geometric measurement of the weld pool, for example, in the measurement of half length of the weld pool.
4. The image quality in the region of part 1 is relatively stable during the welding process.
5. The shape of the weld pool is relative simple and resembles an ellipse. Four parts in different locations are sufficient to locate the boundary contour of the weld pool.
6. Selecting more parts will increase the risk of interference between similar parts and will increase the computation time and decrease the real-time performance of the application.

Figure 3b is the topology of the five-part weld pool model. Without loss of generality, we use the first part as a reference mark, which all the other parts will be measured relative to. The position of all the other parts is assumed to be conditionally independent, given the reference mark. Thus, the model is a tree-structured graphical model of depth 1; there are five vertices and four connections:

$$v = \{v_i(i = 1...5)\},$$

$$e = \{(v_1, v_2), (v_1, v_3), (v_1, v_4), (v_1, v_5)\}$$

which build up a tree. Gaussian distributions are used to model the location of each part relative to the reference mark.



(a) (b)

Fig. 3 Illustration of the part-based model of GTAW pool. In **a**, digital numbers *1–5* denote the part of *top*, *top-left*, *top-right*, *bottom-left*, *bottom-right*, respectively. **b** is the topology of the five-part model

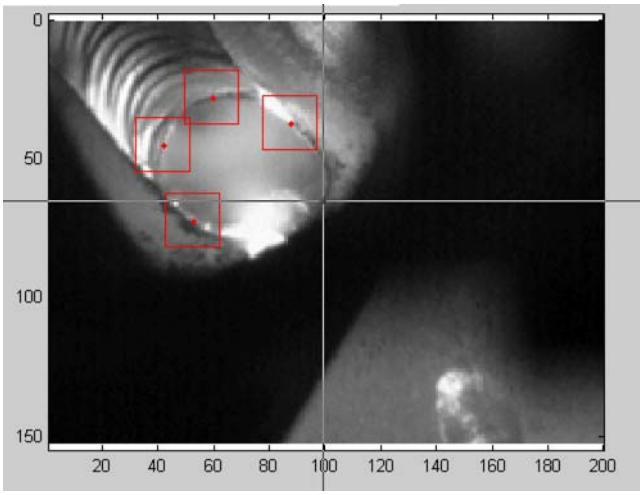


Fig. 4 Illustration of part choice by manual clicks during model training. The *red squares* denote different parts

This model can be easily applied to the geometric parameter measurement of weld pool. Since each part is chosen near the weld pool boundary, the central points of all the parts are nearly coplanar. Thus, some geometric parameters of the weld pool, such as pool width and pool length, are proportional to any edge length of the tree-structured graphical model. The scale factor among them is constant, which can be figured out by simple human–computer interaction during the model training.

Before weld pool detection, we should learn the model parameters first. We will randomly choose a subset of the total image samples for model training, for example, 10% or 20% of the whole samples. In this paper, we manually chose the parts $v_i (i = 1 \ldots 5)$ in the weld pool image. Figure 4 illustrates how to choose the five parts of the weld pool by manual click during model training; see Section 4 for some details. Based on our experiments, we noticed that a click on the area that is seriously spoiled by reflection will distort the model training result. Thus, we should avoid using such weld pool images whose boundaries are seriously spoiled by reflection for model training. The scale of the regions (radius of region) picked out by each click is given in advance, which should be a constant value and is application dependent. In this paper, the radius of each circular region is set as either 10 or 20 pixels of the side length of the square region.

For the appearance of each part $\boldsymbol{v}_i$, the arithmetic mean $\mu_i (i = 1 \ldots 5)$ of the raw pixel intensities of each part over all images is computed as:

$$\mu_i = \frac{1}{N_{\text{train}}} \sum_{j=1}^{N_{\text{train}}} I_{ij}$$

μ_i is taken as the iconic representation of each part, which will act as the part-filter templates in the subsequent weld pool detection. To account for the distribution of the shape, the arithmetic mean and the variance of relative locations are also computed, which will be used as the parameters of the relative location model.

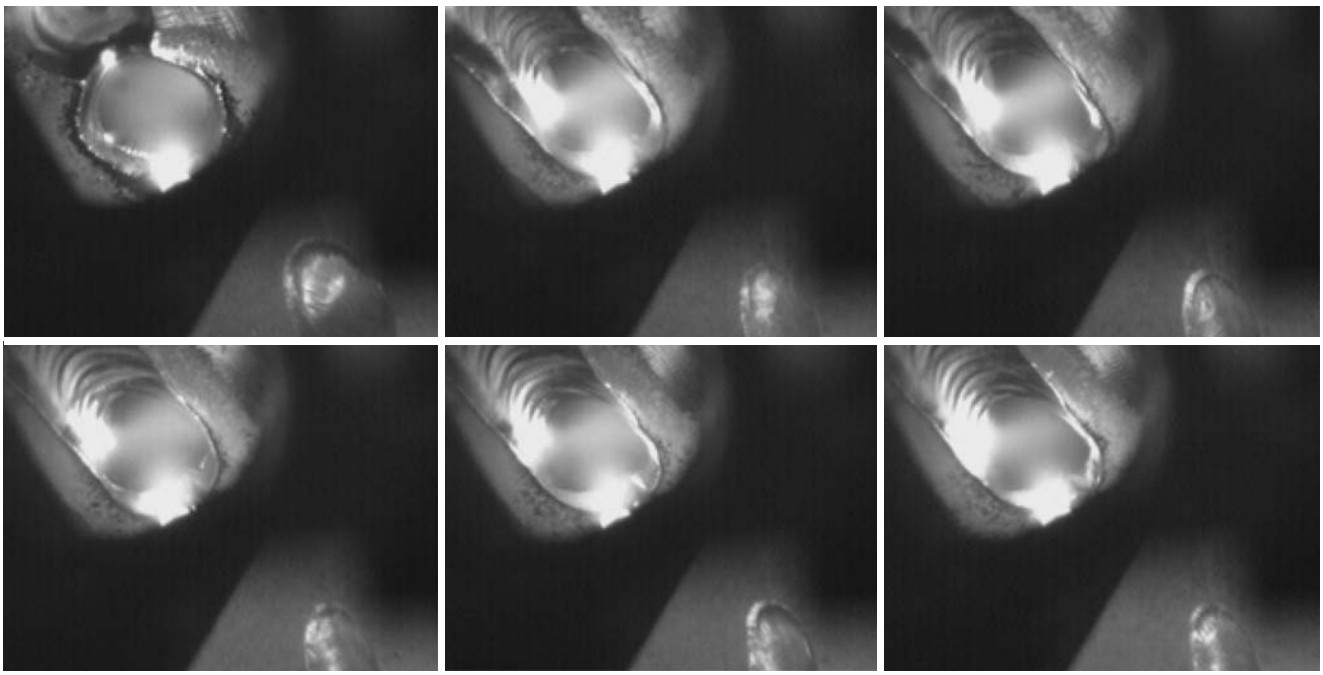


Fig. 5 Some weld pool images captured in a welding process. The *top-left* part of each image is captured from the topside of the weld pool, and the *bottom-right* part is the backside of the weld pool

Note that the learnt model is only translation invariant, not scale or rotation invariant. Fortunately, in practice, the pose of CCD camera relative to the work piece is constant; thus, the shape variance of weld pool during the welding procedure is small and can be neglected.

3.2 Weld pool detection and localization by learnt model

In order to detect and localize the weld pool in the image, we will first run an image part filtering over each image using the trained part-filter templates described above. These templates are used as normalized cross-correlation [16] templates to compute correlation score of image at each point. Normalized cross-correlation is the simplest but very effective similarity measure and is invariant to linear brightness and contrast variations. Its compatibility with hardware implementation such as GPU makes it an ideal choice for real-time applications.

The normalized cross-correlation similarity score between the image block B_{xy} of each points (x, y) and the part filters $\mu_i (i = 1...5)$ are defined as:

$$S = \frac{\sum_{x=-r}^{r}\sum_{y=-r}^{r}\left[\mu_{i,xy} - \overline{\mu_i}\right]\left[B_{xy} - \overline{B}\right]}{(2r+1)^2\sigma(\mu_i)\sigma(B)}, (i = 1...5)$$

where $\overline{\mu_i}(\overline{B})$ is the average, and $\sigma(\mu_i)(\sigma(B))$ is the standard deviation of all the elements in $\mu_i(B)$. For each part filter, we compute the normalized cross-correlation similarity scores over the image and store the locations that achieve the highest similarity score.

Finally, by solving the energy minimization problem of Eq. 5 for each image, we can obtain the best matching configuration of five parts in each image, i.e., the locations of each part will be obtained. The geometric measure of the weld pool can then be easily computed after the location of the weld pool is obtained.

4 Experiment and discussion

In this section, we will conduct some experiments with real images to verify the proposed approach for detection and localization of GTAW pool.

The experimental sensing system we used in this paper is similar to the system used in Wang et al. [7]. The parameters of the light filter system are as follows: the primary filter is a 560–700-nm glass filter; only light with wavelength longer than 560 nm or shorter than 700 nm can pass through it, so it filters out the high and intense noise of the argon, etc. The attenuation of the dimmer glass is 30%. As the light path is composed of topside and backside imaging light paths, we can capture the topside and backside of the weld pool simultaneously during the welding procedure. Figure 5 shows some weld pool images captured during a welding process, where the top-left part of each image is captured from the topside of the weld pool, and it is the region of interest in this paper, and the bottom-right part of each image corresponds to the backside of the weld pool. As we adopt aluminum alloy as workpiece in the test, the reflection of weld pool is less serious. We collected ten groups of images; each group corresponds to an independent welding process and contains about 30~40 images. The image size is 200× 152. For each group of images, we will conduct a model training and a detection/localization test.

The five parts of weld pool are chosen by manual clicks during model training; see Fig. 4. Once an appropriate point for a part is chosen by mouse click, a red square box centering at that point will appear, whose side length is twice that of the predefined radius. Figure 6 demonstrates a model training result of one image group. Figure 6a shows the computed part filter, which will be used as part templates for the later image part filtering step. Figure 6b is the final relative location model of the five parts, where

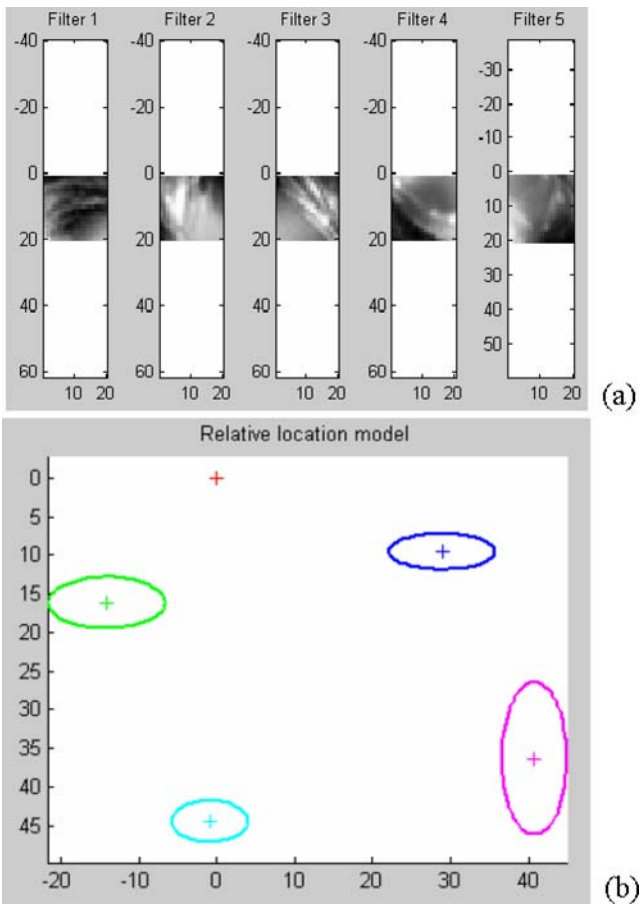


Fig. 6 Model training result of one image group. **a** Five filter parts from model training. **b** Relative location model of five parts, wherein, the top part is the reference mark. The size unit in this figure is pixel

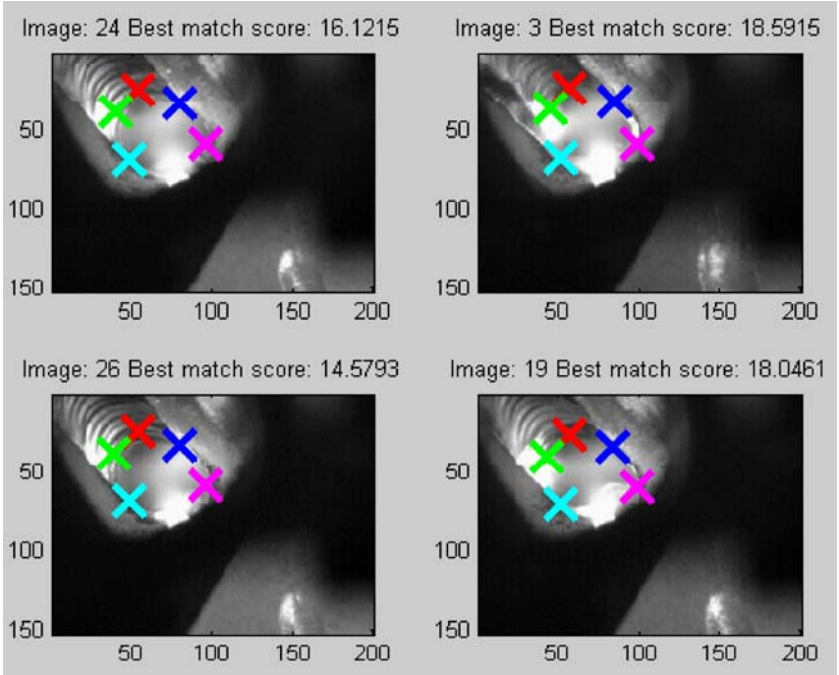


Fig. 7 Some results of visual detection and localization of weld pool. The *crosses with different colors* denote the location of different parts

the "+" denotes location and the corresponding ellipse denotes variance of the part.

Figure 7 shows some detection and localization results of weld pool image, where different colors of crosses denote different parts, along with the best matching score for each image. From the results, we can see that the proposed method is very robust. It works even on weld pool image with strong reflection on its boundary. Moreover, the pileup area around the top part of the weld pool image can also be successfully discerned.

In order to evaluate the performance of localization accuracy, we also conducted some comparative tests using different sizes of subset of the whole image sample during model training. The ground truth locations of the five parts of each weld pool are determined manually. For each image sample, the mean distance between five ground truth locations and the corresponding calculated locations is viewed as localization error. Table 1 shows the result of the comparative tests under different training sample sizes (from 10% to 25% of the total samples). From Table 1, we can see that the localization accuracy continues to improve apparently up to 20% of the total samples for model training. After that, adding more training samples can actually give less improvement.

As to the computation performance of proposed localization method, computing time per image is in the range of 0.4083 s to 0.4762 s, which is reasonable in the application of pulsed GTAW. Since only the top-left area of each image is the region of interest in our application, we can greatly decrease the computation time by only dealing with the top-left area.

5 Conclusion

In this paper, we proposed a robust visual detection and localization approach for the pool of GTAW based on part-based modeling and recognition of objects. This approach provides an elegant framework for representing the outline of a weld pool, especially efficient for weld pool detection and localization in cluttered scenes, where it is partially spoiled or where similar-looking pileup area can act as additional distracters. This approach can provide an initial estimation of the geometric parameter measurement of weld pool. Experiments on real images verified the proposed method.

Acknowledgement This work is supported in part by the National Natural Science Foundation of China (no. 60805018) and Young Faculty Research Grant of Shanghai Jiao Tong University (no. 070110) and Shanghai Sciences & Technology Committee (no. 09JC1407100). The authors wish to thank the anonymous reviewers for their valuable comments on the earlier draft of this paper.

Table 1 Comparative test of localization accuracy under different amounts of training sample

Percent of total samples for model training	10%	15%	20%	25%
Mean localization error (pixels)	2.6940	2.1133	1.8688	1.7537

References

1. Brzakovic D, Khani DT (1991) Weld pool edge detection for automated control of welding. IEEE Trans Robot Autom 7 (3):397–403
2. Kovacevic R, Zhang YM, Ruan S (1995) Sensing and control of weld pool geometry for automated GTA welding. ASME J Eng Indust 117:210–22
3. Chen SB, Lou YJ, Wu L, Zhao DB (2000) Intelligent methodology for sensing, modeling and control of pulsed GTAW: part 1—bead-on-plate welding. Weld J 79(6):151–163
4. Mnich C, Al-Bayat F, Debrunner C, Steele J, Vincent T (2004) In situ weld pool measurement using stereovision. In: ASME Proc. 2004, Japan–USA Symp. on Flexible Automation, Denver, CO, pp 19–21
5. Li LP, Lin T, Chen SB (2005) Light intensity analysis of a passive visual sensing system in GTAW. Int J Adv Manuf Technol 27(1–2):106–111
6. Du QY, Chen SB, Lin T (2006) Inspection of weld shape based on the shape from shading. Int J Adv Manuf Technol 27(7–8):667–671
7. Wang JJ, Lin T, Chen SB (2005) Obtaining weld pool vision information during aluminum alloy TIG welding. Int J Adv Manuf Technol 26(3):219–227
8. Shen HY, Ma HB, Lin T, Chen SB (2007) Research on weld pool control of welding robot with computer vision. Ind Rob 34 (6):467–475
9. Shen HY, Wu J, Lin T, Chen SB (2008) Arc welding robot system with seam tracking and weld pool control based on passive vision. Int J Adv Manuf Technol 39(7–8):669–678
10. Song HS, Zhang YM (2007) Three-dimensional reconstruction of specular surface for a gas tungsten arc weld pool. Meas Sci Technol 18:3751–3767
11. Fischler MA, Elschlager RA (1973) The representation and matching of pictorial structures. IEEE Trans Comput 22(1):67–92
12. Felzenszwalb P, Huttenlocher D (2005) Pictorial structures for object recognition. Int J Comput Vis 61(1):55–79
13. Leibe B, Leonardis A, Schiele B (2008) Robust object detection with interleaved categorization and segmentation. Int J Comput Vis 77(1–3):259–289
14. Carneiro G, Lowe DG (2006) Sparse flexible models of local features. European Conference on Computer Vision (ECCV), Graz, Austria, Part III, pp 29–43
15. Rao RPN, Ballard DH (1995) An active vision architecture based on iconic representations. Artif Intell 78(1/2):461–505
16. Zhang Z, Deriche R, Faugeras O, Luong QT (1995) A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artif Intell 78(1–2):87–119