Title:
Small Visible Defect Detection based on Small Sample Analysis

Authors:
Zhichao Sun, sunzhichao@sjtu.edu.cn, Shanghai Jiao Tong University
Xiangzhi Wei, antonwei@sjtu.edu.cn, Shanghai Jiao Tong University

Introduction:
Small visible defects often appear on the surface of parts during the manufacturing process. Visible defects directly affect the appearance of parts and the sales. Therefore, it is necessary to ensure the surface quality of parts, especially for daily used stuff (e.g., the surface of a watch or jewelry). Machine vision is a common and effective detection method for this purpose. At present, visual detection based on deep learning has made a great breakthrough in terms of accuracy [3],[8]. However, this method is not suitable for the production of small-scaled products, since it relies on a large amount of data to become effective, and the training of the deep learning model usually requires a very long time [6]. Typically, the production of a new design is limited to a small scale, and the initial sample is not sufficient for the detection function to predict small visible defects. In this study, we shall focus on the detection of the small visible defects. Note that there are visible defects and invisible defects on the surface of a dial (Fig. 1), and invisible defects do not affect the appearance and the sale of the product, therefore they are allowed in qualified products. Note that our method is also applicable to the other products.
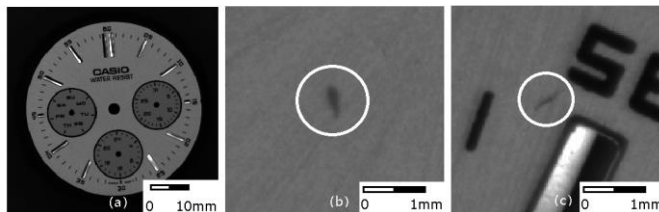

Fig. 1: Defects on a dial: (a) a dial, (b) visible defects, (c) invisible defects.

Main Idea
To address this issue, we develop a fast approach for small visible defect detection based on the fusion of multiple features using a small sample of data. Given an initial evaluation function based on a fairly small sample of data, we developed a recursive scheme to tune the coefficients of the evaluation function as the sample size increases during the production process. In other words, the evaluation function is updated by feeding it with new data as the production goes on. We shall show that this method can effectively prevent over-fitting, improve the prediction accuracy and enhance the robustness of the prediction model. To validate the proposed approach and illustrate our method, a test experiment using a patch of 500 watch dials (Fig. 1) was conducted, and the experimental results show that it can identify visible defects efficiently.

Preprocessing

The role of preprocessing is to extract the ROI (Region Of Interest, including small visible defect regions and invisible defect regions) of the dial image. Given a set of dial images and a standard dial image, we shall preprocess them as follows: The standard mask of the standard image is obtained by edge detection. Then the image matching based on pyramid search strategy is used to match the standard mask and the dial image to mask the irrelevant regions of dial image, and then ROI of dial image is obtained by mean filtering and image difference.

Canny edge detection [2],[7] can be used to find the boundaries of the standard mask. Edge detection starts with image smoothing by a Gaussian filter. The Gaussian function is as shown in Eqn. (1.1).

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\frac{x^2 + y^2}{2\sigma^2} \tag{1.1}$$

And then we applied a non-maximum suppression to the magnitude of the gradient of the standard image to determine the edge pixels. The magnitude of the gradient can be computed by Eqn. (1.2).

$$M(x, y) = \sqrt{g_x^2 + g_y^2} \tag{1.2}$$

Where g_x and g_y are expressed in Eqn. (1.3).

$$\begin{cases} g_x = \left[f_s(x+1, y) - f_s(x, y) + f_s(x+1, y+1) - f_s(x, y+1) \right] / 2 \\ g_y = \left[f_s(x, y+1) - f_s(x, y) + f_s(x+1, y+1) - f_s(x+1, y) \right] / 2 \end{cases} \tag{1.3}$$

Thereafter, we conducted a double-threshold processing and connectivity analysis. As a result, a standard mask can be obtained, which can be used to mask the irrelevant regions of the dial images. To reduce the match time, we applied a pyramid search strategy as follows:

The pyramid data structure consists of 4 images. The lowest level of the pyramid is the original image. Subsequently, the image of the upper layer is obtained from a lower layer by low-pass filtering, and its resolution is lower than the image in the lower layer.

When matching using the pyramid data structure, the standard mask is matched in a top-to-down fashion. Finally, the exact position of the standard mask on the dial is obtained. In this process, the Hausdorff distance is used to measure the matching rate [4].

Refer to Fig. 2 for an example, the dial images of each layer and the matching results are shown. From left to right, the image resolution and the matching accuracy become higher.
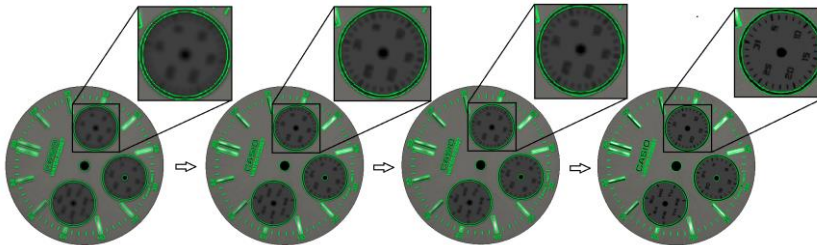

Fig. 2: The process of matching, the green part of each image is the standard mask (transformed w.r.t the background image).

As the transformation result of the standard mask relative to the current image is determined, the standard mask can be used to mask the irrelevant regions of the image. And then the mean filtering is applied to the image. Finally, the ROI can be extracted from the difference of image before and after the mean filtering, the calculation formula is as Eqn. (1.4).

$$D(x, y) = | g(x, y) - f(x, y) | \tag{1.4}$$

Where $f(x, y)$ is the image before mean filtering, $g(x, y)$ is the image after mean filtering, and $D(x, y)$ is ROI.

Statistical Analysis

Multiple regression analysis (MRA) is performed on the regions extracted (by using preprocessing to extract the region) from 40 images (the images are from the initial sample image before the production process) to establish the evaluation function. The preliminarily selected variables are $F_i \ i \ = 1,2,3,4,5$, which are defined as follows: Area F_1 is the area of a defect; Convexity $F_2 = F_0 / F_c$, where F_c is the convex hulled area of the region, and F_0 is the area of the region; Rectangularity $F_3 = A / A_r$, where A is the area of the region, and A_r is the area of the minimum enclosing rectangle of the region; length-width $F_4 = L_l / L_w$, where L_l and L_w are the length and width of the minimum enclosing rectangle of the region, respectively. Compactness $F_5 = max\{1, e^2 / 4\pi F_1\}$, where e is the peripheral length of the defect region.

Next we use the stepwise regression method to further screen the variables [1],[5]. The basic idea of stepwise regression is to add the variables one by one. We use the F-test (p<0.1) to judge whether a variable is significant.

With the above processing, only F_1 , F_2 , F_1^2 and F_2^2 are significant, other variables are not significant, and the evaluation function can be expressed in Eqn. (1.5).

$$Y = 0.02527F_1 - 17F_2 - 0.000103F_1^2 + 9.82F_2^2 + 6.88 \tag{1.5}$$

In the equation, where Y is the evaluation value. If the response value $Y > 0.5$, the region is considered to be a visible defect, else otherwise.

In addition, the accuracy of the evaluation function can be evaluated by Eqn. (1.6).

$$R^2 = 1 - \frac{\sum (Y\_actual - Y\_predict)^2}{\sum (Y\_actual - Y\_mean)^2} \tag{1.6}$$

A high value of R^2 indicates a strong correlation between the variables and the value of Y .

Recursive Tuning of the Coefficients of the Evaluation Function

The original evaluation function can be further optimized by tuning the coefficients using the new data obtained from the production process. More precisely, in Eqn. (1.5), the coefficient $k_1 = 0.02527$, $k_2 = -17$, $k_3 = -0.000103$, $k_4 = 9.82$, and $k_5 = 6.88$ need to be tuned.

According to each region *r* extracted (by using preprocessing to extract the region) from 50 images (the images were taken from the production process), calculate its correction value $\Delta k_m^{(r)}(m = 1,2,3,4,5)$ corresponding to $k_m(m = 1,2,3,4,5)$ (in Eqn. 1.5) by following steps:

To evaluate the difference between the evaluated value (Eqn. 1.5) and the actual value (1 or 0, visible or not, observed by the users), we define the loss function in Eqn. (1.7).

$$Cost^{(r)} = \frac{1}{2}[\sigma(Y^{(r)}) - Y^{(r)*}]^2 \tag{1.7}$$

where $Y^{(r)}$ is calculated from Eqn. (1.5), $Y^{(r)*}$ is the actual value of region r . $\sigma(x)$ is the activation function for transforming the value of $Y^{(r)}$ into a range of [0, 1] (since the value of $Y^{(r)*}$ is either 0 or 1), and can be expressed in Eqn. (1.8).

$$\sigma(x) = \frac{1}{1 + e^{-x+0.5}} \tag{1.8}$$

By taking the partial derivative of the loss function, the correction value $\Delta k_m^{(r)}(m = 1,2,3,4,5)$ of the value $k_m^{(r)}$ can be calculated in Eqn. (1.9).

$$\Delta k_m^{(r)} = -\eta \frac{\partial Cost}{\partial k_m} = -\eta \frac{\partial Cost}{\partial \sigma(Y)} \cdot \frac{\partial \sigma(Y)}{\partial Y} \cdot \frac{\partial Y}{\partial k_m} = -\eta(\sigma(Y) - Y^*) \cdot \sigma'(Y) \cdot V_k \tag{1.9}$$

where η is the step size, V_k is the values of variables.

The correction value $\Delta k_m^{(r)}(m = 1,2,3,4,5)$ is calculated from a single region r. However, the correction value of a single region cannot reflect the characteristics of all regions, and therefore we use the average of the correction values of all regions for this purpose. More precisely, we calculate the actual correction value $\Delta k_m(m = 1,2,3,4,5)$, which is the average of the correction values of all regions extracted from 50 images, and its formula is shown in Eqn. (1.10).

$$\Delta k_m = \frac{\sum_{r=1}^{n} \Delta k_m^{(r)}}{n} \tag{1.10}$$

Therefore, the value of k_m in Eqn. (1.5) is updated by Eqn. (1.11):

$$k_m = k_m + \Delta k_m \tag{1.11}$$

If the change of loss between two iterations is less than a user-defined threshold value, then the iteration is terminated. Through 12 iterations based on 50 images, the updated evaluation function is obtained in Eqn (1.12):

$$Y = 0.0252717F_1 - 17.00F_2 + 0.000159397F_1^2 + 9.82F_2^2 + 6.88 \tag{1.12}$$

In order to verify the method, we use 500 dials and 500 small jewelry accessories to do the test. The optimized evaluation function is used to predict the type of defects of the ROI (obtained by preprocessing). The prediction accuracy of dials and jewelry accessories is 95.8% and 94.8%. Fig. 3 and Fig. 4 shows example images and their prediction results. It can be seen from the result that our proposed algorithm can detect the defects successfully. Note that the products are cheap and therefore the accuracy level is acceptable, and that the accuracy can be further improved by using more data as the production goes on.

Our method has no fixed time of accomplishing a full cycle. It takes less time to generate the evaluation function in the early stage, the overall time of one full cycle mainly depends on the time of evaluation function optimization: The longer the optimization process is, the higher the detection accuracy will be. In terms of the programming skills required by the user, our algorithm can be realized easily since it only involves a few critical math formulas (e.g., Eqn. (1.5) and Eqn. (1.12)).
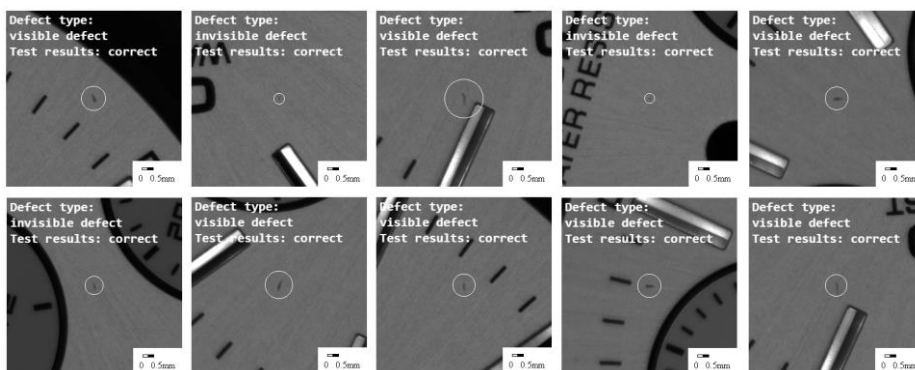
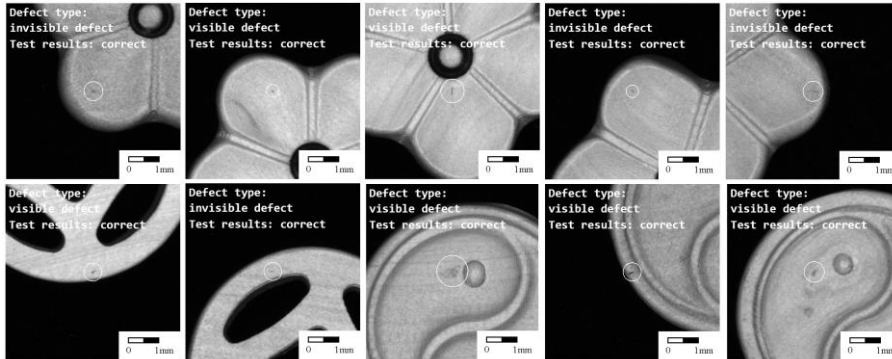Fig. 3: Results of example defect images of dial.

Fig. 4: Results of example defect images of jewelry accessory.

Conclusions:
Fast visible defect detection is of critical importance for screening the unqualified products. For a small-scaled production, the evaluation function is usually inappropriate due to the lack of data. To address this issue, we proposed a scheme of tuning the coefficients of the evaluation functions based on the continuing tested parts. The experimental results show that this method can continuously optimize the detection model with the small samples, and overcome the low detection accuracy of the primary model. However, the initial evaluation of the truth still requires the intervention of users. How to automatically determine the truth is worthy of future research.

References:
[1]   Ahmed, A. B.; Musonda, I.; Pretorius, J. H.: Natural logarithm transformation for predicting procurement time of ppp projects in nigeria, Cogent Engineering, 6(1), 2019. https://doi.org/10.1080/23311916.2019.1571147
[2]   Canny, J.F.: A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine, 8(6), 1986, 679 698. https://doi.org/10.1109/TPAMI.1986.4767851
[3]   Chen, F. C.; Jahanshahi, R. M. R.: Nb-cnn: deep learning-based crack detection using convolutional neural network and naïve bayes data fusion, IEEE Transactions on Industrial Electronics, 65(99), 2018, 4392-4400. https://doi.org/10.1109/TIE.2017.2764844
[4]   Lalonde; Marc; Beaulieu; Mario; Gagnon; Langis: Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching, IEEE transactions on medical imaging, 20(11), 2001. https://doi.org/10.1109/42.963823
[5]   Maestro-Watson, D.; Balzategui, J.; Eciolaza, L.; Arana-Arexolaleiba, N.: Deflectometric Data Segmentation Based on Fully Convolutional Neural Networks, International Society for Optics and Photonics: Mulhouse, SPIE 11172, 2019, 1117209. https://doi.org/10.1117/12.2521740
[6]   Sun, X.; Gu, J.; Tang, S.; Li, J.: Research progress of visual inspection technology of steel products —a review, Applied Sciences, 8(11), 2018, 2195. https://doi.org/10.3390/app8112195
[7]   Wang, H.; Nie, D.; Tuo, X.; Zhong, Y.: Research on crack monitoring at the trailing edge of landslides based on image processing, Landslides, 17(4), 2020, 985-1007. https://doi.org/10.1007/s10346-019-01335-z
[8]   Wei, X.; Yang, Z.; Liu, Y.; Wei, D.; Jia, L.; Li, Y.: Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study, Engineering Applications of Artificial Intelligence, 80(APR.), 2019, 66–81. https://doi.org/10.1016/j.engappai.2019.01.008