

(19) **DANMARK**

(10) **DK/EP 4209993 T3**



(12)

Oversættelse af  
europæisk patentskrift

Patent- og  
Varemærkestyrelsen

- 
- (51) Int.Cl.: **G 06 T 7/00 (2017.01)**
- (45) Oversættelsen bekendtgjort den: **2024-10-28**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2024-10-02**
- (86) Europæisk ansøgning nr.: **21958589.0**
- (86) Europæisk indleveringsdag: **2021-10-26**
- (87) Den europæiske ansøgnings publiceringsdag: **2023-07-12**
- (86) International ansøgning nr.: **CN2021126393**
- (87) Internationalt publikationsnr.: **WO2023070312**
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **Contemporary Amperex Technology (Hong Kong) Limited, Level 19, China Building, 29 Queen's Road Central, Central, Central And Western District, Hong Kong**
- (72) Opfinder: **JIANG, Guannan, , Fujian 352100, Kina**  
**HAN, Lili, , Fujian 352100, Kina**
- (74) Fuldmægtig i Danmark: **RWS Group, Europa House, Chiltern Park, Chiltern Hill, Chalfont St Peter, Bucks SL9 9FG, Storbritannien**
- (54) Benævnelse: **FREMGANGSMÅDE TIL BILLEDBEHANDLING**
- (56) Fremdragne publikationer:  
**CN-A- 102 508 259**  
**CN-A- 103 337 069**  
**CN-A- 103 930 923**  
**CN-A- 112 541 886**  
**CN-U- 209 432 721**  
**JUAN LI ET AL: "FUSION OF LIDAR 3D POINTS CLOUD WITH 2D DIGITAL CAMERA IMAGE", A THESIS, 1 January 2015 (2015-01-01), XP055722335, Retrieved from the Internet**  
**<URL:http://www.secs.oakland.edu/~li4/research/student/JuanLi2015.pdf> [retrieved on 20200813]**  
**GUAN LIMIN ET AL: "Real-Time Vehicle Detection Framework Based on the Fusion of LiDAR and Camera", ELECTRONICS, vol. 9, no. 3, 7 March 2020 (2020-03-07), pages 451, XP093094214, DOI: 10.3390/electronics9030451**



# DESCRIPTION

Description

## Technical Field

[0001] The present application relates to the field of image processing, and in particular, to an image processing method as specified in claims 1 to 6 and a non-volatile storage medium specified in claim 7.

## Background Art

[0002] During industrial generation, a product may have a certain defect due to process and equipment reasons. Therefore, it is necessary to detect whether the product is defective, to improve the yield rate of products. In conventional defect detection methods, manual detection has disadvantages such as low detection efficiency. By means of visual image technologies, the detection efficiency may be improved, and also the cost of human input may be greatly reduced. Current visual detection methods usually employ a CCD camera to obtain two-dimensional (2D) images, which are then used for detection. However, such a detection method using 2D images is not accurate enough, which is likely to lead to missed and false detection of the defects. Juan Li ET AL: "FUSION OF LIDAR 3D POINTS CLOUD WITH 2D DIGITAL CAMERA IMAGE", A thesis, 1 January 2015 (2015-01-01), XP055722335, mentions fusion data of multi-sensors. Guan Limin ET AL: "Real-Time Vehicle Detection Framework Based on the Fusion of LiDAR and Camera", Electronics, vol.9, no.3, 7 March 2020 (2020-03-07), page 451, XP093094214, DOI: 3390/electronics9030451, mentions a vehicle detection by using LiDAR and Camera. CN209432721U mentions a lithium battery surface defects detection equipment.

## Summary of the Invention

[0003] In view of the problems described above, the present application proposes an image processing method as specified in claims 1 to 6 and a non-volatile storage medium as specified in claim 7.

[0004] To this end, a first aspect of the present invention provides an image processing method, where the image processing method includes:

obtaining a two-dimensional image and a three-dimensional image of an object, wherein the object is selected from any one of the following: a battery surface, a battery top cover weld, and a battery sealing pin;

extracting luminance information and depth information from the three-dimensional image to generate a luminance map and a depth map for the object;

graying the two-dimensional image to obtain a single-channel grayscale image;

selecting at least same three location points from each of the grayscale image and the luminance map; calculating a coordinate transformation matrix between the grayscale image and the luminance map based on coordinates of the at least same three location points in each of the grayscale image and the luminance map; and

aligning the two-dimensional image with the luminance map and the depth map based on the coordinate transformation matrix, to obtain fused image data, where the fused image data includes color information of the two-dimensional image and the luminance information and depth information of the three-dimensional image, wherein the fused image data of the object is input into an updated deep learning algorithm model to obtain a defect determining result of the object.

**[0005]** In the present invention, the two-dimensional image is grayed to obtain the single-channel grayscale image. The coordinate transformation matrix between the two-dimensional image and the three-dimensional image is then solved. Features of the two-dimensional image are aligned with those of the three-dimensional image, such that feature data of the two-dimensional image is aligned with that of the three-dimensional image, so as to obtain the fused image data. In this way, image fusion of the two-dimensional image with the three-dimensional image is implemented. Such fused image data may provide more information for subsequent deep learning model analysis, thereby obtaining a more accurate detection result.

**[0006]** The purpose of graying is to facilitate alignment of the features in the two-dimensional image with those in the three-dimensional image. In some embodiments, a weighted average method may be used to gray a color image with three (RGB) channels. Specifically, a grayscale value may be calculated according to the following formula:

$$p = w_1R + w_2G + w_3B$$

where  $p$  represents a grayscale value of a coordinate point,  $R$  represents a red value of the coordinate point,  $w_1$  represents a weight of the red value,  $G$  represents a green value of the coordinate point,  $w_2$  represents a weight of the green value,  $B$  represents a blue value of the coordinate point, and  $w_3$  represents a weight of the blue value. Research shows that human eyes are most sensitive to green, followed by red, and are least sensitive to blue. Therefore, in some embodiments,  $w_2 > w_1 > w_3$  may be set. In some embodiments,  $w_1$  may be set to 0.299,  $w_2$  may be set to 0.587, and  $w_3$  may be set to 0.114. This is because experiments and theories proof that using such weight settings may obtain psychological grayscales that best match

perception of human eyes. Red, green, and blue information of the image are fully preserved in grayscale values obtained by using the weighted average method. It should be understood that graying may be performed using other methods, such as a component method, a maximum method, an average method and a Gamma correction method.

**[0007]** In some embodiments, the at least same three location points are points of the object with characteristics. These points with characteristics are particular locations in patterns that are within the image and may be obtained by analysis of the patterns. Therefore, selection of the at least same three location points may be implemented using artificial intelligence algorithms.

**[0008]** In some embodiments, the at least same three location points include three non-collinear reference points. When the three reference points are collinear, the coordinate transformation matrix cannot be obtained from coordinates of the three reference points. Therefore, the at least same three location points are required to include three non-collinear reference points.

**[0009]** In some preferred embodiments, a maximum interior angle of a triangle formed by the three reference points is not greater than  $160^\circ$ . The larger the maximum interior angle of the triangle formed by the three reference points is, the more the three reference points tend to be collinear. Small sampling errors may cause the three reference points to tend to be collinear, such that the coordinate transformation matrix obtained based on the coordinates of the three reference points in each of the grayscale image and the luminance map is ineffective. Therefore, the maximum interior angle of the triangle formed by the three reference points is set to not greater than  $160^\circ$ , which improves the reliability of the coordinate transformation matrix.

**[0010]** In some specific embodiments, the image processing method includes:

aligning, the luminance map and the depth map with the two-dimensional image based on the coordinate transformation matrix by using coordinates in the two-dimensional image as a reference; and

using, as information of three channels respectively, grayscale information of the grayscale image,

and the luminance information and the depth information that are aligned to the coordinates in the two-dimensional image, to generate a fused image containing the information of the three channels.

**[0011]** It should be understood that one embodiment of the present application is to align the luminance map and the depth map obtained from the three-dimensional image with the two-dimensional image by using the coordinates in the two-dimensional image as a reference. It

can be conceived by those skilled in the art that aligning the two-dimensional image with the luminance map, that is, aligning the two-dimensional image to the three-dimensional image by using coordinates in the luminance map as a reference implements alignment of the two-dimensional image with the three-dimensional image, and further completes data alignment of the two images. However, the resolution of the three-dimensional image is low, so the luminance map and the depth map are preferably aligned to the two-dimensional image. Moreover, it should be understood that it is only necessary to align information of the two-dimensional image with that of the three-dimensional image, and store color information, luminance information, and depth information, etc. for each coordinate in the image, rather than synthesizing the fused image containing the information of the three channels.

**[0012]** In some embodiments, the color information, the luminance information, and the depth information may be channel stacked to obtain an image having information of multiple channels. In some other embodiments, the color information, the luminance information, and the depth information may be fused by weighting to obtain an image having information of a single channel or multiple channels.

**[0013]** In some specific embodiments, before the two-dimensional image is aligned with the luminance map and the depth map, depth values in the depth map are normalized to generate a normalized depth map. Depth values in the generated depth map may be very large, e.g., ten thousand or above, thus by means of normalization, the depth values may be limited between 0 and 1, which facilitates subsequent calculation and processing. In some embodiments, the depth values in the depth map may be normalized according to the following formula:

$$d_s = \frac{d - d_{min}}{d_{max} - d_{min}}$$

where  $J$  represents a depth value of a coordinate point in the depth map,  $d_{min}$  represents a minimum depth value in the depth map,  $d_{max}$  represents a maximum depth value in the depth map, and  $d_s$  represents a normalized depth value of the coordinate point in the depth map.

**[0014]** In the present invention, the object described above is selected from any one of the following: a battery surface, a battery top cover weld, and a battery sealing pin. The embodiments of the present application are applicable to a scenario where both a two-dimensional image and a three-dimensional image are required for defect detection. During battery production, detection procedures of scratches on the battery surface, defects of the battery top cover weld, and defects of the sealing pin weld are all important for ensuring safety of a battery after leaving the factory. The two-dimensional image cannot directly reflect depth information of related parts of the battery, and therefore effective imaging of depth-sensitive features cannot be implemented, while the three-dimensional image has low resolution and cannot cover all target features. Therefore, the scratches on the battery surface, the defects of the battery top cover weld, and the defects of the sealing pin weld may be detected by using the image processing method according to the present application, so as to obtain fused image data for corresponding feature detection.

**[0015]** In some embodiments, the two-dimensional image and the three-dimensional image described above may be images acquired at different positions of a production line. However, in some other embodiments, the two-dimensional image and the three-dimensional image described above may be images acquired at the same position of a production line.

**[0016]** In some embodiments, the two-dimensional image and the three-dimensional image described above may be images of the same object taken at a consistent angle. However, in some other embodiments, the two-dimensional image and the three-dimensional image described above may be images of the same object taken at different angles.

**[0017]** Another aspect of the present application provides a non-volatile storage medium storing instructions, where the instructions, when executed by a processor, cause the processor to perform the image processing method of the first aspect described above.

**[0018]** Another embodiment of the present application provides an appearance detection method, including:

obtaining fused image data of a plurality of objects according to the image processing method of the first aspect described above;

obtaining defect determining results of the plurality of objects;

using the fused image data and the defect determining results of the plurality of objects to iteratively train a deep learning algorithm model, to update the deep learning algorithm model;

acquiring a two-dimensional image and a three-dimensional image of a target object;

fusing the two-dimensional image and the three-dimensional image of the target object according to the image processing method of the first aspect described above, to obtain fused image data of the target object; and, as defined in the first aspect, inputting the fused image data of the target object into an updated deep learning algorithm model, to obtain a defect determining result of the target object.

**[0019]** In embodiments of the present application, the defect determining result of the target object is obtained by substituting the fused image data of the target object into the trained model. Thus, the accuracy of the detection is improved and missed and false detection (false positive) of defects are reduced. At the same time, since the fused image data is substituted into one model for calculation and analysis, there is no need to construct two models or respectively substitute data of different images into the two models for calculation and analysis, thereby saving computing resources and increasing a detection speed. Moreover, since the fused image data is used, only one model rather than two models need to be trained in advance.

**[0020]** In some embodiments, the fused image data of the target object includes grayscale

information, luminance information, and depth information of the target object, and the deep learning algorithm model described above includes a grayscale data input channel, a luminance data input channel, and a depth data input channel. In the embodiments of the present application, the deep learning algorithm model may include three data input channels, and the model detects, based on input grayscale information, luminance information, and depth information, whether there is a defect.

**[0021]** In some embodiments, the deep learning algorithm model described above includes a red data input channel, a green data input channel, a blue data input channel, a luminance data input channel, and a depth data input channel. In the embodiments of the present application, the deep learning algorithm model may include five data input channels, and the model detects, based on input color information, luminance information, and depth information, whether there is a defect. It should be understood that, in practice, those skilled in the art may choose the number of data channels of the model according to actual requirements.

**[0022]** In some embodiments, the acquired two-dimensional image of the target object includes multiple two-dimensional images.

**[0023]** In some embodiments, the acquired three-dimensional image of the target object includes multiple three-dimensional images.

**[0024]** Another embodiment of the present application provides a non-volatile storage medium storing instructions, where the instructions, when executed by a processor, cause the processor to perform the appearance detection method.

### **Brief Description of the Drawings**

**[0025]** In order to illustrate the technical solutions of the present application more clearly, the drawings required in the description of the embodiments of the present application will be described briefly below. Obviously, the drawings described below are merely some embodiments of the present application, and for those of ordinary skill in the art, other drawings can also be obtained from these drawings without any creative efforts. In the drawings:

FIG. 1 illustrates a flowchart of an image processing method according to an embodiment of the present application;

FIG. 2 illustrates a flowchart of an image processing method according to an embodiment of the present application;

FIGs. 3a to 3d show pictures of a two-dimensional image, a luminance map, a depth map and a fused image according to an embodiment of the present application; and

FIG. 4 illustrates a flowchart of an appearance detection method according to an embodiment

of the present application.

### **Detailed Description of Embodiments**

**[0026]** Embodiments of the present application will be described in more detail below with reference to the drawings. The following embodiments are merely intended to more clearly illustrate the technical solutions of the present application, so they merely serve as examples, but are not intended to limit the scope of protection of the present application.

**[0027]** Unless otherwise defined, all technical and scientific terms used herein have the same meanings as those commonly understood by those skilled in the art to which the present application belongs. The terms used herein are merely for the purpose of describing specific implementations, but are not intended to limit the present application. The terms "comprising" and "having" and any variations thereof in the description and the claims of the present application as well as the brief description of the accompanying drawings described above are intended to cover non-exclusive inclusion.

**[0028]** The phrase "embodiment" mentioned herein means that the specific features, structures, or characteristics described in conjunction with the embodiment can be encompassed in at least one embodiment of the present application. The phrase at various locations in the description does not necessarily refer to the same embodiment, or an independent or alternative embodiment exclusive of another embodiment. Those skilled in the art understand explicitly or implicitly that the embodiment described herein may be combined with another embodiment.

**[0029]** If steps are stated in sequence in the description or in the claims, this does not necessarily mean an embodiment or aspect is limited to the stated sequence. Conversely, it may be envisaged that the steps are performed in a different sequence or in parallel to each other, unless one step is established on another, which absolutely requires the established step to be performed subsequently (this will be clear in an individual case). Therefore, the stated sequence may be a preferred embodiment.

**[0030]** The inventors of the present application have found in practice that since a 2D image cannot not directly reflect depth information of each part of a product and cannot fully cover defect forms in the product, current methods for detecting product defects using a 2D image easily leads to missed and false detection of defects. Moreover, the inventors have also recognized that a 3D image acquired by using a three-dimensional (3D) camera has low resolution, and also cannot fully cover defect forms in the product, which easily leads to missed and false detection of defects. The inventors have conceived that, since the 2D image has enough resolution, it may be used to preliminarily detect product defects, and then the 3D image is used to re-detect the product defects, to eliminate false detection (false positive) and

add missed defects.

**[0031]** Still further, the inventors have realized that, if features of the 2D image and the 3D image are extracted and respectively substituted into two different neural network models for analysis, two times of analysis need to be performed, and the two neural network models needs to be trained in advance. On this basis, the inventors have conceived to fuse the 2D image with the 3D image to obtain fused image data and substitute the data into a signal neural network model for analysis. Therefore, the required computing resources are reduced, and the analysis speed and the efficiency are improved.

**[0032]** FIG. 1 illustrates a flowchart of an image processing method 100 according to an implementation aspect of the present application. As shown in FIG. 1, in step 102, a two-dimensional image and a three-dimensional image of an object are obtained. In step 104, luminance information and depth information are then extracted from the three-dimensional image to generate a luminance map and a depth map. In step 106, the two-dimensional image is grayed to obtain a single-channel grayscale image, to facilitate alignment of features in the two-dimensional image with those in the three-dimensional image. In step 108, after the luminance map and the grayscale image are obtained, at least same three location points are selected from each of the grayscale image and the luminance map. In step 110, a coordinate transformation matrix between the grayscale image and the luminance map is calculated based on coordinates of the at least same three location points in each of the grayscale image and the luminance map.

**[0033]** It should be understood by those skilled in the art that the same feature point in images acquired by using two different cameras has different image point coordinates. Since product forms in images taken by using different cameras have the same flatness and parallelism, a coordinate transformation between image point coordinates of the different cameras is an affine transformation, with the following matrix transformation equation:

$$\begin{bmatrix} qx \\ qy \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} px \\ py \\ 1 \end{bmatrix}$$

where  $(px, py)$  represents image point coordinates of one camera,  $(qx, qy)$  represents image point coordinates of the other camera, and  $a_1, b_1, c_1, a_2, b_2,$  and  $c_2$  represent transformation coefficients in the coordinate transformation matrix. Such affine transformation has six transformation coefficients, i.e., six degrees of freedom. Therefore, three groups of points are required to calculate the transformation coefficients in the coordinate transformation matrix.

**[0034]** In step 112, after the coordinate transformation matrix is calculated, the two-dimensional image is aligned with the luminance map and the depth map based on the coordinate transformation matrix, to obtain fused image data. In the present invention, the two-dimensional image is grayed to obtain the single-channel grayscale image. The coordinate transformation matrix between the two-dimensional image and the three-dimensional image is then solved. Features of the two-dimensional image are aligned with those of the three-dimensional image, such that feature data of the two-dimensional image is aligned with that of the three-dimensional image, so as to obtain the fused image data. In this way, image fusion of

the two-dimensional image with the three-dimensional image is implemented. Such fused image data may provide more information for subsequent deep learning model analysis, thereby obtaining a more accurate detection result.

**[0035]** FIG. 2 illustrates a flowchart of an image processing method according to an embodiment of the present application. As shown in FIG. 2, in step 202, a three-dimensional image of an object is obtained. In step 206, luminance information and depth information are extracted from the three-dimensional image to generate a luminance map and a depth map, as shown in FIGs. 3b and 3c. Then, in step 210, depth values in the depth map are normalized to generate a normalized depth map. In some embodiments, the depth values in the depth map may be normalized according to the following formula:

$$d_s = \frac{d - d_{min}}{d_{max} - d_{min}}$$

where  $J$  represents a depth value of a coordinate point in the depth map,  $d_{min}$  represents a minimum depth value in the depth map,  $d_{max}$  represents a maximum depth value in the depth map, and  $d_s$  represents a normalized depth value of the coordinate point in the depth map. Depth values in the depth map generated in step 206 may be very large, e.g., ten thousand or above, thus by means of normalization, the depth values may be limited between 0 and 1, which facilitates subsequent calculation and processing.

**[0036]** In step 204, a two-dimensional image of the object, as shown in FIG. 3a, is obtained. Then, in step 208, the two-dimensional color image with three (RGB) channels is grayed by weighting to obtain a single-channel grayscale image. Specifically, a grayscale value may be calculated according to the following formula:

$$p = w_1R + w_2G + w_3B$$

where  $p$  represents a grayscale value of a coordinate point,  $R$  represents a red value of the coordinate point,  $w_1$  represents a weight of the red value,  $G$  represents a green value of the coordinate point,  $w_2$  represents a weight of the green value,  $B$  represents a blue value of the coordinate point, and  $w_3$  represents a weight of the blue value. Research shows that human eyes are most sensitive to green, followed by red, and are least sensitive to blue. Therefore, in some embodiments,  $w_2 > w_1 > w_3$  may be set. In some embodiments,  $w_1$  may be set to 0.299,  $w_2$  may be set to 0.587, and  $w_3$  may be set to 0.114. This is because experiments and theories prove that using such weight settings may obtain psychological grayscales that best match perception of human eyes. Obtaining the grayscale value by using a weighted average method is easily implemented using algorithms with a relatively small calculation amount. In addition, red, green, and blue information of the image are fully preserved. It should be understood that graying may be performed using other methods, such as a component method, a maximum method, an average method and a gamma correction method.

**[0037]** It should be understood that, as shown in FIG. 2, step 202, step 206 and step 210 may be performed before, after, or when step 204 and step 208 are performed. In step 212, after steps 206 and 208 are completed, at least same three location points with characteristics may be selected from each of the grayscale image and the luminance map. These points with

characteristics are particular locations in patterns that are within the image and may be obtained by analysis of the patterns. Therefore, selection of the same three location points with characteristics may be implemented using artificial intelligence algorithms. FIGs. 3a to 3d illustrate an embodiment of welding defect detection of a lithium battery sealing pin, according to the present application. In this embodiment, in the figures, two central points of angles between a straight weld and a circular weld, and the center point on the sealing pin may be selected as the same three location points with characteristics. The same three location points are non-collinear, so as to meet calculation requirements of transformation coefficients in a coordinate transformation matrix.

**[0038]** In some preferred embodiments, a maximum interior angle of a triangle formed by the same three location points is not greater than  $160^\circ$ . The larger the maximum interior angle of the triangle formed by the same three location points is, the more three reference points tend to be collinear. Small sampling errors may cause the same three location points to tend to be collinear, such that the coordinate transformation matrix obtained based on the coordinates of the same three location points in each of the grayscale image and the luminance map is ineffective. Therefore, the maximum interior angle of the triangle formed by the same three location points is set to not greater than  $160^\circ$ , which improves the reliability of the coordinate transformation matrix. When the same three location points do not meet the requirement that the points are non-collinear or the requirement that the maximum interior angle of the formed triangle is not greater than  $160^\circ$ , more location points may be selected, as long as three reference points from the selected same location points meet either of the requirements above.

**[0039]** In step 214, a coordinate transformation matrix between the grayscale image and the luminance map is calculated based on coordinates of the same three location points in each of the grayscale image and the luminance map. Step 214 is similar to step 110 in FIG. 1, and related content will not be repeated herein.

**[0040]** In step 216, after the coordinate transformation matrix is calculated, the luminance map and the normalized depth map are aligned to the two-dimensional image based on the coordinate transformation matrix. In this way, data alignment of the same location points can be implemented in the dimension of pixels of the two-dimensional image, and thus, the effectiveness and accuracy of information fusion is ensured.

**[0041]** Finally, in step 218, the coordinate transformed luminance map and depth map are channel stacked with the grayscale image, to form a fused image having information of three channels, as shown in FIG. 3d. Moreover, it should be understood that it is only necessary to align information of the two-dimensional image with that of the three-dimensional image in coordinates, and store color information, luminance information, and depth information, etc. for each coordinate in the image, rather than synthesizing the fused image containing the information of the three channels. In some other embodiments, the color information, the luminance information, and the depth information may alternatively be fused by weighting to obtain an image having information of a single channel or multiple channels.

**[0042]** The image processing method of the present application is applicable to a scenario where both a two-dimensional image and a three-dimensional image are required for defect detection. During battery production, detection procedures of scratches on the battery surface, defects of the battery top cover weld, and defects of the sealing pin weld are all important for ensuring safety of a battery after leaving the factory. The two-dimensional image cannot directly reflect depth information of related parts of the battery, and therefore effective imaging of depth-sensitive features cannot be implemented, while the three-dimensional image has low resolution and cannot cover all target features. Therefore, the scratches on the battery surface, the defects of the battery top cover weld, and the defects of the sealing pin weld may be detected by using the image processing method according to the present application, so as to obtain fused image data for corresponding feature detection. FIGs. 3a to 3d illustrate an embodiment of welding defect detection of a lithium battery sealing pin, according to the present application. However, it should be understood that the object in the image processing method of the present application may alternatively be a battery surface, a battery top cover weld, etc. In the embodiment illustrated in FIGs. 3a to 3d, the two-dimensional image in FIG. 3a and the three-dimensional image corresponding to FIGs. 3b and 3c are images acquired at different positions of a production line. However, in some other embodiments, the two-dimensional image and the three-dimensional image may be images acquired at the same position of a production line.

**[0043]** FIG. 4 illustrates a flowchart of an appearance detection method according to the present application. As shown in FIG. 4, in step 402, fused image data of a plurality of objects are obtained according to the image processing method according to the present application. In step 404, defect determining results of the plurality of objects are obtained. The defect determining results obtained in step 404 are determined by visual inspectors. Then, in step 406, the fused image data and the defect determining results of the plurality of objects are used to iteratively train a deep learning algorithm model, update the deep learning algorithm model. In step 408, after a useful deep learning algorithm model is obtained through training, a two-dimensional image and a three-dimensional image of a target object are acquired. In step 410, the two-dimensional image and the three-dimensional image of the target object are fused according to the image processing method as described before, to obtain fused image data of the target object. Finally, in step 412, the fused image data of the target object is input into an updated deep learning algorithm model to obtain a defect determining result of the target object.

**[0044]** In the embodiment illustrated in FIG. 4, the defect determining result of the target object is obtained by substituting the fused image data of the target object into the trained model. Thus, the accuracy of the detection is improved and missed and false detection (false positive) of defects are reduced. At the same time, since the fused image data is substituted into one model for calculation and analysis, there is no need to construct two models or respectively substitute data of different images into the two models for calculation and analysis, thereby saving computing resources and increasing a detection speed. Moreover, since the fused image data is used, only one model rather than two models need to be trained in advance.

**[0045]** In some embodiments, the fused image data of the target object described above includes grayscale information, luminance information, and depth information of the target object, and the deep learning algorithm model described above includes a grayscale data input channel, a luminance data input channel, and a depth data input channel. In this embodiment, the deep learning algorithm model may include three data input channels, and the model detects, based on input grayscale information, luminance information, and depth information, whether there is a defect.

**[0046]** In some embodiments, the deep learning algorithm model described above includes a red data input channel, a green data input channel, a blue data input channel, a luminance data input channel, and a depth data input channel. In this embodiment, the deep learning algorithm model may include five data input channels, and the model detects, based on input color information, luminance information, and depth information, whether there is a defect. It should be understood that, in practice, those skilled in the art may choose the number of data channels of the model according to actual requirements.

**[0047]** The present invention is not limited to the specific embodiments disclosed herein but includes all the technical solutions that fall within the scope of the appended claims.

## **REFERENCES CITED IN THE DESCRIPTION**

Cited references

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

**Patent documents cited in the description**

- [CN209432721U \[0002\]](#)

**Non-patent literature cited in the description**

- **JUAN LI et al.** FUSION OF LIDAR 3D POINTS CLOUD WITH 2D DIGITAL CAMERA IMAGEA thesis, 2015, [0002]
- **GUAN LIMIN et al.** Real-Time Vehicle Detection Framwork Based on the Fusion of LiDAR and CameraElectronics, 2020, vol. 9, 3451- [0002]

## Patentkrav

1. Fremgangsmåde til computer-implementeret billedbehandling, der omfatter:

5 opnåelse af et todimensionelt billede og et tredimensionelt billede af et objekt (102; 408), hvor objektet vælges fra et hvilket som helst af følgende: en batterioverflade, en svejsning på batteriets topdæksel og en batteriforseglingsstift;  
udtrækning af luminansinformation og dybdeinformation fra det  
10 tredimensionelle billede med henblik på at generere et luminanskort og et dybdekort for objektet (104);  
gråtoning af det todimensionelle billede for at opnå et enkeltkanals gråtonebillede (106);  
valg af mindst samme tre positionspunkter fra hvert enkelt af  
15 gråtonebilledet og luminanskortet (108, 212);  
beregning af en koordinattransformationsmatrix mellem gråtonebilledet og luminanskortet baseret på koordinater for de mindst samme tre positionspunkter i hvert enkelt af gråtonebilledet og luminanskortet (110, 214); og  
20 tilpasning af det todimensionelle billede til luminanskortet og dybdekortet baseret på koordinattransformationsmatrixen for at opnå sammenlagte billeddata (112; 410), hvor de sammenlagte billeddata omfatter farveinformation om det todimensionelle billede og luminansinformationen og dybdeinformationen i det  
25 tredimensionelle billede,  
hvor de sammenlagte billeddata for objektet indlæses i en indlært dyb læringsalgoritmemodel for at opnå et fejlbestemmelsesresultat for objektet (412).

30 2. Fremgangsmåde til billedbehandling ifølge krav 1, hvor de mindst samme tre positionspunkter er punkter for objektet med karakteristikker (212).

35 3. Fremgangsmåde til billedbehandling ifølge krav 1 eller 2, hvor de mindst samme tre positionspunkter omfatter tre ikke-kollinære referencepunkter.

4. Fremgangsmåde til billedbehandling ifølge krav 3, hvor en maksimal indvendig vinkel af en trekant, der er dannet af de tre referencepunkter ikke er større end  $160^\circ$ .

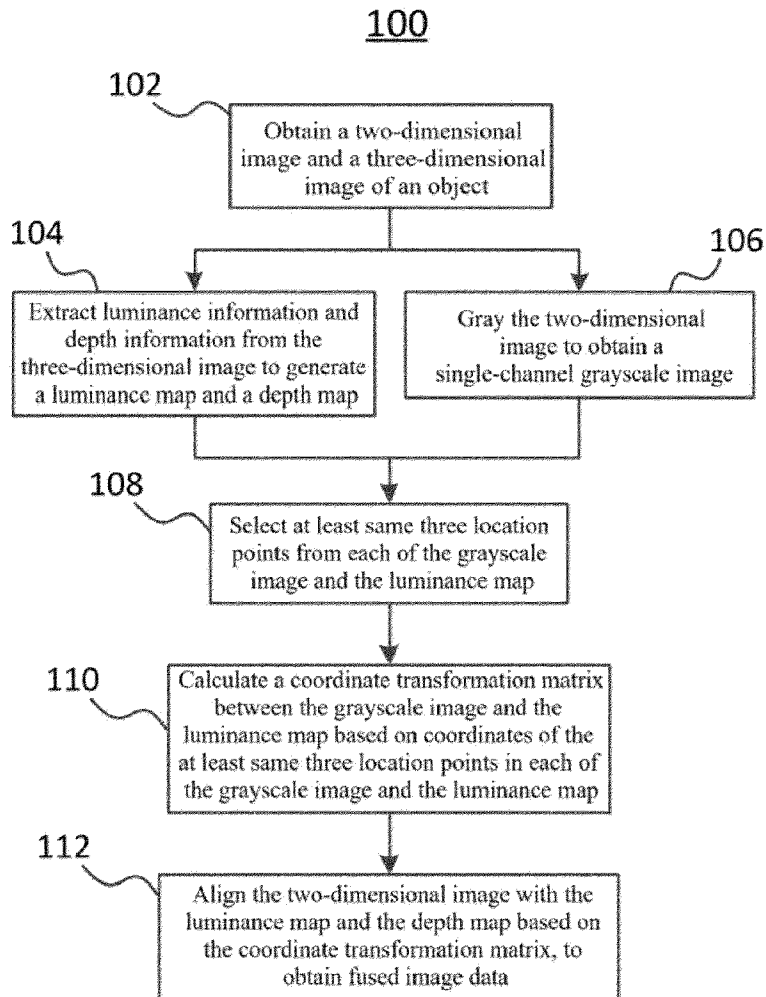
5 5. Fremgangsmåde til billedbehandling ifølge et hvilket som helst af kravene 1 til 4, der omfatter:  
tilpasning af luminanskortet og dybdekortet til det todimensionelle billede baseret på koordinattransformationsmatrixen ved at bruge koordinater i det  
10 todimensionelle billede som reference; og  
anvendelse, som information fra tre respektive kanaler, af gråtoneinformation fra gråtonebilledet og luminansinformationen og dybdeinformationen, der er tilpasset koordinaterne i det  
15 sammenlagt billede, der indeholder informationen fra de tre kanaler.

6. Fremgangsmåde til billedbehandling ifølge et hvilket som helst af kravene 1 til 5, hvor dybdeværdier i dybdekortet  
20 normaliseres for at generere et normaliseret dybdekort (210), inden det todimensionelle billede tilpasses til luminanskortet og dybdekortet.

7. Ikke-flygtigt lagermedium, der lagrer instrukser, der, når  
25 de udføres af en processor, får processoren til at gennemføre fremgangsmåden ifølge et hvilket som helst af kravene 1 til 6.

# DRAWINGS

Drawing



*FIG. 1*

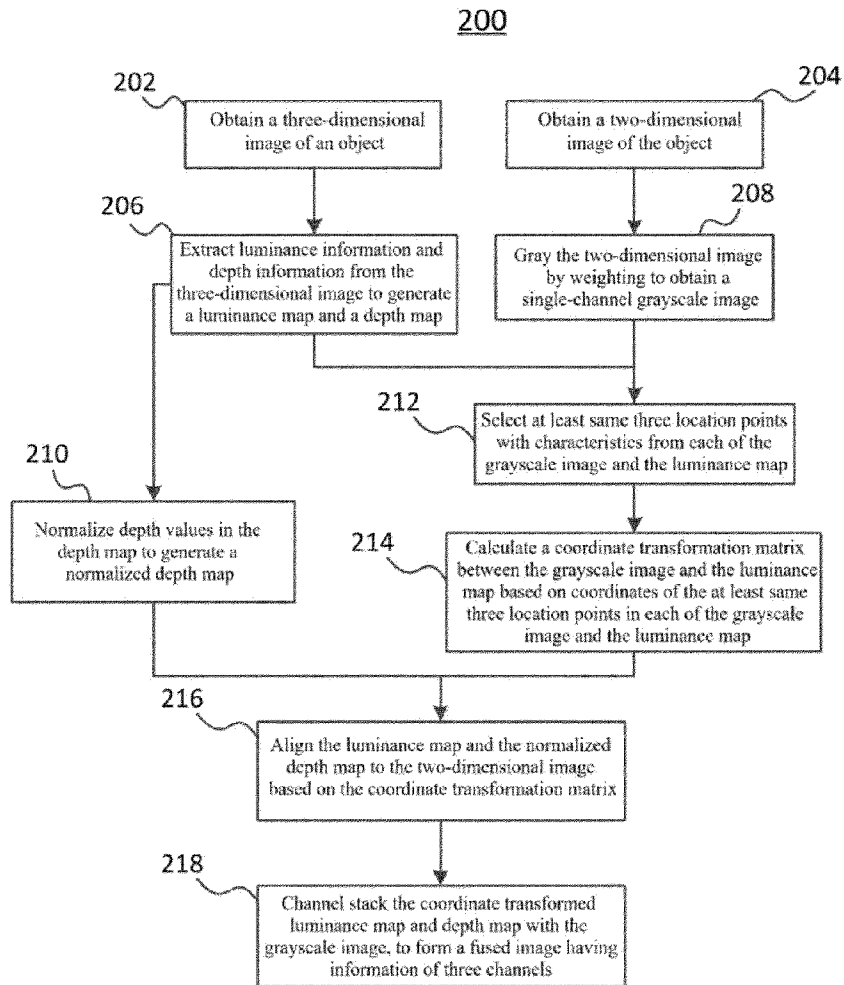


FIG. 2

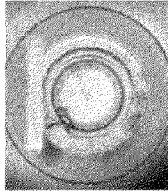


FIG. 3a

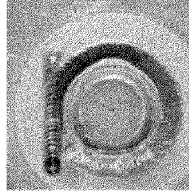


FIG. 3b

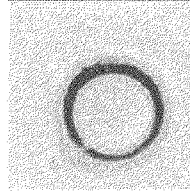


FIG. 3c

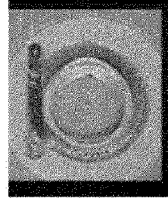


FIG. 3d

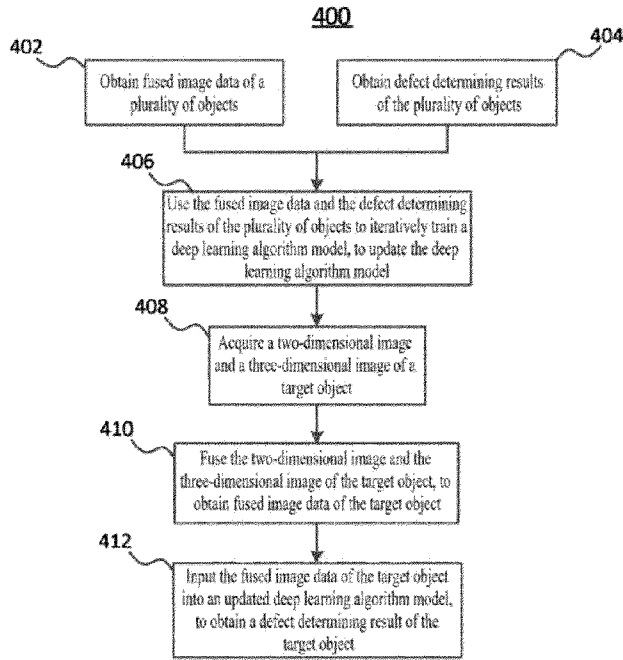


FIG. 4