Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3094424 C 2022/02/22

(11)(21) 3 094 424

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2020/03/30
(87) Date publication PCT/PCT Publication Date: 2020/12/21
(45) Date de délivrance/Issue Date: 2022/02/22
(85) Entrée phase nationale/National Entry: 2020/09/24
(86) N° demande PCT/PCT Application No.: CN 2020/082006
(87) N° publication PCT/PCT Publication No.: 2020/253308
(30) Priorité/Priority: 2019/06/21 (CN2019105403497)

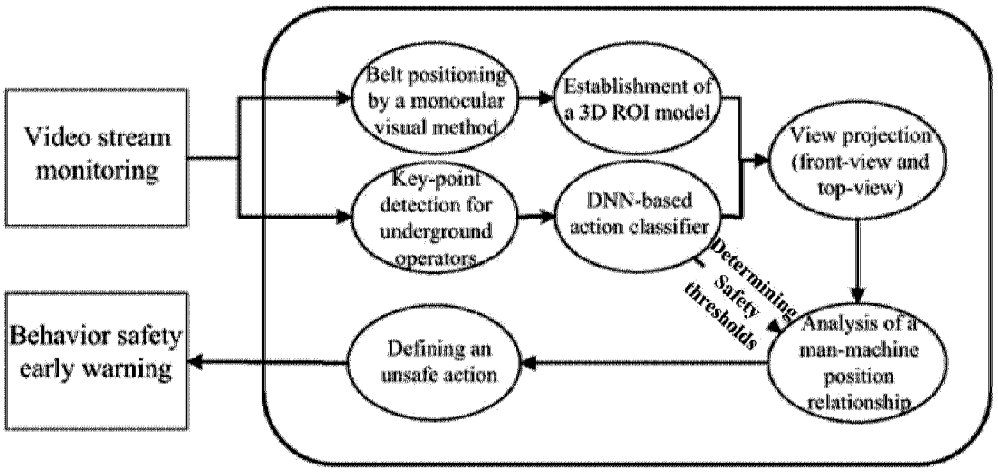
(51) Cl.Int./Int.Cl. *G08B 21/02* (2006.01), *B65G 43/00* (2006.01), *G06N 3/02* (2006.01), *H04N 7/18* (2006.01)

(72) Inventeurs/Inventors:
SUN, YANJING, CN;
DONG, KAIWEN, CN;
CHENG, XIAOZHOU, CN;
YUN, XIAO, CN;
HOU, XIAOFENG, CN;
WANG, BOWEN, CN;
...

(73) Propriétaire/Owner:

(54) Titre : PROCEDE D'AVERTISSEMENT PRECOCE ET DE SURVEILLANCE DE SECURITE POUR LE COMPORTEMENT D'INTERACTION HOMME-MACHINE D'UN OPERATEUR DE BANDE TRANSPORTEUSE SOUTERRAINE

(54) Title: SAFETY MONITORING AND EARLY-WARNING METHOD FOR MAN-MACHINE INTERACTION BEHAVIOR OF UNDERGROUND CONVEYOR BELT OPERATOR



(57) Abrégé/Abstract:
A safety monitoring and early-warning method for a man-machine interaction behavior of an underground conveyor belt operator is disclosed. The present invention determines a belt position in a video based on a camera calibration principle, and delineates a three-dimensional region of interest (ROI) based on the belt position and size. By using a "bottom-up" key-point extraction method, multi-person key-point detection of conveyor belt operators is performed by detection followed by clustering, thus guaranteeing the detection accuracy and improving the detection efficiency. The human body key-points and the ROI are separately projected twice, and a relationship between the human body and the belt position is estimated based on the two projections to screen out unsafe actions during man-machine interaction and give an early warning, so as to avoid a major safety hazard in the belt conveyor system caused by an abnormal contact of the operator with the belt region.

CA 3094424 C 2022/02/22
(11)(21) 3 094 424
(13) C

(72) Inventeurs(suite)/Inventors(continued): WANG, BIN, CN; XU, HONGLI, CN; CHEN, XIAOJING, CN
(73) Propriétaires(suite)/Owners(continued):CHINA UNIVERSITY OF MINING AND TECHNOLOGY, CN
(74) Agent: BLANEY MCMURTRY LLP

ABSTRACT

A safety monitoring and early-warning method for a man-machine interaction behavior of an underground conveyor belt operator is disclosed. The present invention determines a belt position in a video based on a camera calibration principle, and delineates a three-dimensional region of interest (ROI) based on the belt position and size. By using a "bottom-up" key-point extraction method, multi-person key-point detection of conveyor belt operators is performed by detection followed by clustering, thus guaranteeing the detection accuracy and improving the detection efficiency. The human body key-points and the ROI are separately projected twice, and a relationship between the human body and the belt position is estimated based on the two projections to screen out unsafe actions during man-machine interaction and give an early warning, so as to avoid a major safety hazard in the belt conveyor system caused by an abnormal contact of the operator with the belt region.

Date Reçue/Date Received 2020-09-24

SAFETY MONITORING AND EARLY-WARNING METHOD FOR MAN-MACHINE INTERACTION BEHAVIOR OF UNDERGROUND CONVEYOR BELT OPERATOR

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to the field of monitoring of underground operations, and in particular, to a behavior safety monitoring method for an underground conveyor belt operator.

Description of Related Art

China leads the world in the development of coal production industry all along. However, as a high-risk industry, coal mining has posed great production safety hazards over the years. Belt conveyors in coal mines, as the most common underground transport system currently, directly affect the safety level of coal production during operation. Safety management of the belt conveyor system mainly resorts to a manual monitoring manner at the present stage, which has many limitations such as short operation duration, a narrow functional coverage, and a high cost. Therefore, it is of great significance to develop a safety early-warning system based on video monitoring for the belt conveyors and related operators, thus improving the production safety level of the belt conveyor system.

The current behavior safety early-warning system based on video monitoring pre-warns coal miners of danger merely by analyzing and identifying actions of the miners. For example, a behavior safety monitoring method based on feature extraction and support vector machine (SVM) classification that was proposed by Yang chaoyu et al. in 2016, and an underground dangerous region monitoring method based on moving target detection that was proposed by Zhang liya in 2017 both use rectangles to locate underground operators, implementing behavior safety monitoring of the operators. For limitations of the methods using the rectangles, Zhu aichun et al. in 2018 proposed a method for identifying postures of underground coal miners based on an hourglass

Date Reçue/Date Received 2020-09-24

network with hard example mining for generative adversarial training, which implements positioning and safety identification of underground operators by detecting human body key-points, improving accuracy and robustness of safety identification for the underground operators. The foregoing methods achieve good evaluation and identification effects for unsafe behaviors without man-machine interaction (namely, interaction between human and equipment). However, most underground accidents occur in a process of unsafe man-machine interaction. Therefore, without identification of man-machine interaction behaviors, safety monitoring and early warning cannot be accurately implemented merely by identifying actions or determining positions of the operators. Moreover, an existing algorithm model (for example, the hard example mining for generative adversarial training used by Zhu aichun et al.) is complex in structure and slow in operation speed, and has a problem that the detection speed linearly decreases as the number of detected persons increases, failing to have a wide application prospect.

SUMMARY OF THE INVENTION

Technical Solution

To solve the technical problems mentioned above in the prior art, the present invention provides a safety monitoring and early-warning method for a man-machine interaction behavior of an underground conveyor belt operator.

To achieve the foregoing technical objective, the present invention adopts the following technical solutions:

A safety monitoring and early-warning method for a man-machine interaction behavior of an underground conveyor belt operator includes the following steps:

(1) acquiring an underground real-time video stream by using a surveillance camera;

(2) estimating the size of a belt in the video according to a camera calibration principle, and accordingly delineating a three-dimensional region of interest (ROI), namely, a belt dangerous region;

(3) detecting human body key-points of all persons in the video, measuring the

Date Reçue/Date Received 2021-05-26

degree of correlation between the key-points by using part affinity fields, and clustering key-points belonging to each individual by means of matching and optimization in a bipartite graph, thus achieving the objective of detecting human body key-points of each person in the video;

(4) determining x-axis and y-axis components of each detected human body key-point in a world coordinate system, self-defining a height component z for the human body key-point, and then combining the three components into complete world coordinates of the human body key-point; and

(5) determining, according to a relative position relationship between the belt dangerous region and the human body key-points of each person, whether a man-machine interaction behavior is safe, thus determining whether to give an early warning.

Further, in step (3), with each picture in the video as an input, deep features thereof are extracted to obtain a feature graph F, and then the feature graph F is input to step 1 having two convolutional neural network (CNN) branches; in this step, the first CNN branch predicts confidence maps $S^1=\rho^1(F)$ regarding a set of key-points, where ρ^1 indicates an inference procedure of this CNN branch in step 1; the second CNN branch predicts a set of "part affinity fields" $\mathbf{L}^1= \emptyset^1(\mathbf{F})$, where $\emptyset^1$ indicates an inference procedure of this CNN branch in step 1 and aims to cluster the predicted human body key-points on respective limbs of each individual, to obtain a set of complete information of human body key-points; and afterwards, the prediction results obtained by step 1 based on the two CNN branches are linked in series with the original feature graph F, and then are together input to subsequent steps, so as to obtain a more accurate prediction result, where the subsequent steps are denoted by the following formulas:

$$\mathbf{S}^t = \rho^t(\mathbf{F},\mathbf{S}^{t-1},\mathbf{L}^{t-1}), \quad \forall t \geq 2$$
$$\mathbf{L}^t = \emptyset^t(\mathbf{F},\mathbf{S}^{t-1},\mathbf{L}^{t-1}), \quad \forall t \geq 2$$

where S^t and L^t respectively indicate confidence maps and part affinity fields obtained by step t, and ρ^t and $\emptyset^t$ indicate inference procedures of the two CNN branches in step t respectively.

Further, a loss function, namely, a mean squared error, is applied for the two CNN

Date Reçue/Date Received 2020-09-24

branches separately after each step, and the loss functions for the two CNN branches in step t are respectively as follows:

$$f_{\mathbf{S}}^{t} = \sum_{j=1}^{J} \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \left\| \mathbf{S}_{j}^{t}(\mathbf{p}) - \mathbf{S}_{j}^{*}(\mathbf{p}) \right\|_{2}^{2}$$

$$f_{\mathbf{L}}^{t} = \sum_{c=1}^{C} \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \left\| \mathbf{L}_{c}^{t}(\mathbf{p}) - \mathbf{L}_{c}^{*}(\mathbf{p}) \right\|_{2}^{2}$$

where f_s^t and f_L^t are respectively the loss functions for the two CNN branches in step t; p is the coordinates of any position in a picture to be detected; W(p) is a Boolean value, which equals 0 when there is no mark in a training dataset, or otherwise equals 1; $\mathbf{S}^t_j(\mathbf{p})$ indicates a confidence map of the jth human body key-point at the position p in step t; $\mathbf{S}^*_j(\mathbf{p})$ indicates a true position of the confidence map; $\mathbf{L}^t_c(\mathbf{p})$ indicates a part affinity field at the position p in step t, and $\mathbf{L}^*_c(\mathbf{p})$ indicates a true position of the part affinity field;

a true reference for defining confidence levels of key-points at any position p in the picture is as follows:

$$\mathbf{S}_{j,k}^{*}(\mathbf{p}) = \exp\left(-\frac{\left\| \mathbf{p} - \mathbf{x}_{j,k} \right\|_{2}^{2}}{\sigma^{2}} \right)$$

where $x_{j,k}$ indicates true coordinates of the jth human body key-point of the kth person in a marked training sample, and σ indicates a constant for controlling the degree of dispersion of a Gaussian graph regarding confidence points; and

a maximum value is taken, to obtain a confidence reference $\mathbf{S}_j(\mathbf{p}) = \max S^*_{j,k}(\mathbf{p})$ of the jth human body key-point of the kth person.

Further, each person has 9 human body key-points in total, which are the nose, chest, right shoulder, right hand, left shoulder, left hand, hip bone, right foot, and left foot that represent a person.

Further, in step (5), front-view and top-view projections of the belt dangerous region are determined based on the belt dangerous region obtained in step (2); for each

Date Reçue/Date Received 2020-09-24

person in the video, a minimum distance d_T from the human body key-points to the top-view projection of the belt dangerous region, a minimum distance d_F from the human body key-points to the front-view projection of the belt dangerous region, and the height h of the corresponding human body key-point are calculated; and if d_T and d_F are both less than or equal to a safe distance threshold d and h is less than the height of the front-view projection of the belt dangerous region, it is determined that a man-machine interaction behavior of the operator is unsafe and an alarm is given.

Further, in step (5), a Deep Neural Network (DNN) classifier is established and used to classify actions according to the detected information of the human body key-points; the key-point information in each picture is combined into a sample which corresponds to one action type; a large number of marked human body key-point - action samples are used to train the classifier, such that the classifier has the ability to identify human actions in each single picture; and safe distance thresholds d_i corresponding to different actions are determined according to identification results of the classifier, where the subscript i represents the ith action type.

Further, considering continuity of actions of the operator in a surveillance video, a probability judgment model for multiple successive pictures is added based on the action identification for each single picture: by using M successive pictures as a unit for judgment, returning results of classifying actions in the M pictures one by one with the action classifier, and performing counting for different classification results separately; and finally, calculating a ratio of the counted number in each result to the total number in all the results, and determining the greatest ratio as a classification result of the actions in the M pictures.

Further, there are three action classification results: falling over, squatting down, and smoking; different safety coefficients γ_i are assigned to the three action types respectively, and their respective safe distance thresholds are calculated as follows: $d_i = \gamma_i \cdot d$, where $i = 1, 2, 3$; and finally, it is determined, according to the safe distance threshold, whether man-machine interaction of the operator for the current action is safe.

Date Reçue/Date Received 2020-09-24

Advantageous Effect

The foregoing technical solutions achieve the following advantageous effects:

The present invention determines a belt position in a video based on a camera calibration principle, and delineates a three-dimensional ROI based on the belt position and size. By using a "bottom-up" key-point extraction method, multi-person key-point detection of conveyor belt operators is performed by detection followed by clustering, thus guaranteeing the detection accuracy and improving the detection efficiency. The human body key-points and the ROI are separately projected twice: front-view projection and top-view projection; and a relationship between the human body and the belt position is estimated based on the two projections. A DNN classifier is established and used to classify actions according to the key-point information in each single picture, and action labels of each person are returned. By combining action identification and position judgment, the positions of actions with different unsafe coefficients are determined based on different scales. The present invention screens out unsafe actions during man-machine interaction and gives an early warning, so as to avoid a major safety hazard in the belt conveyor system caused by an abnormal contact of the operator with the belt region.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is an overall flowchart of the present invention;

FIG. 2 is a projection diagram of a belt dangerous region from three angles of view;

FIG. 3 is a schematic diagram of camera calibration;

FIG. 4 is a schematic structural diagram of a key-point predication network;

FIG. 5 is a curve chart showing a relationship between key-point coordinates and confidence levels;

FIG. 6 is a schematic diagram showing transformation of belt coordinates and projection;

FIG. 7 is a schematic simplified diagram of human body key-points;

Date Reçue/Date Received 2020-09-24

FIG. 8 is a schematic diagram of evaluating an unsafe action by a projection method;

FIG. 9 is a schematic diagram of classification of actions of an underground operator;

FIG. 10 is a schematic diagram of safety judgment when the operator falls over;

FIG. 11 is a schematic diagram of safety judgment when the operator squats down;

FIG. 12 is a schematic diagram of safety judgment when the operator smokes; and

FIG. 13 is a schematic diagram of a specific implementation process of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The technical solutions of the present invention are described in detail below with reference to the accompanying drawings.

A procedure of a safety monitoring and early-warning method for a man-machine interaction behavior of an underground conveyor belt operator that is provided by the present invention is shown in FIG. 1. For a real-time video stream acquired by a surveillance camera, a three-dimensional ROI model is established for a belt position based on a camera calibration principle. Then, key-points of the conveyor belt operator are detected based on a "bottom-up" approach and actions are classified based on key-point information by using a DNN. Finally, the ROI and the key-points are projected in the front-view and top-view directions, and a positional relationship between the key-points and the ROI is evaluated according to safe distance thresholds corresponding to different actions, to determine a dangerous action and give an early warning.

1. Modeling for a belt dangerous region

During belt safety identification for early warning, delineating an unsafe ROI on the belt is a basic task in a detection phase. The present invention identifies a dangerous action by evaluating a positional relationship between human body key-points and the delineated ROI on the belt. Because a positional relationship between the human being and the belt in the vertical direction cannot be evaluated by means of a 2D ROI, a false

Date Reçue/Date Received 2020-09-24

alarm rate inevitably increases if the conventional method is used to delineate the 2D ROI on the belt. For example, when a miner normally works on a platform higher than the belt, it is highly probable that such an operation is evaluated as an unsafe behavior according to the 2D ROI. To solve the problem, the present invention proposes establishing a 3D ROI model according to the belt position, estimating the size of the belt in a video according to a camera calibration and imaging principle, and accordingly delineating a 3D ROI region, where a diagram of the ROI from three angles of view is shown in FIG. 2.

2. Measurement of a belt size by camera calibration

(i) A measurement principle of a belt size is as follows: Based on known internal parameters of a monocular camera and coordinates of a picture-image coordinate system in a monocular lens, a relationship between an image coordinate system and a world coordinate system is determined, so as to establish a three-dimensional model showing a relative position of the belt and surrounding operators.

(ii) The image coordinate system is a coordinate system using a pixel as a unit, with the origin being on the upper left side. A position of each pixel point is denoted with pixels, and therefore such a coordinate system is called an image-pixel coordinate system denoted as (u, v), where u and v respectively indicate the column and the row of a pixel point in a digital image.

(iii) The world coordinate system is a three-dimensional coordinate system defined by a user, and used to describe a relative position of an object and the camera in a three-dimensional space, which is denoted as (X, Y, Z).

It can be learned from Figure 3 that the image coordinate system UO_1P, a camera coordinate system with O_2 as the origin, and the world coordinate system XO_3Y are on the upper left corner, where the already known quantities are as follows:

The height of the camera is H; a distance between a world coordinate point corresponding to a pixel coordinate center and the camera on the y axis is O_3M; image coordinates of the pixel coordinate center O_1 are (ucenter, vcenter); a measured point P is a projection of a point Q to be measured on the Y axis of the world coordinate system,

Date Reçue/Date Received 2021-05-26

which has pixel coordinates $P_1(0, v)$; an actual pixel length is x_{pix} and an actual pixel width is y_{pix}; and O_1O_2 indicates a focal length f of the camera. A schematic diagram of calibration is shown in FIG. 3.

A coordinate of Y is calculated as follows:

$$\alpha = \arctan\left(\frac{H}{O_3M}\right),$$

$$\gamma = \arctan\left(\frac{O_1P_1 \times ypix}{f}\right) = \frac{(v - vcenter) \times ypix}{f},$$

$\beta = \alpha - \gamma$,

$$O_3P = \frac{H}{\tan(\beta)},$$

where γ indicates an included angle between O_1O_2 and P_1P, and α is an angle between the camera and the horizontal plane and is denoted by an acute angle formed between O_1O_2 and the Y axis. After the angle β is obtained by calculation, the coordinate $Y=O_3P$ in the vertical direction can be calculated according to the properties of a right triangle.

A coordinate of X is calculated as follows:

$$O_2P_1 = \sqrt{[(v - vcenter) \times xpix]^2 + f^2},$$

$$O_2P = \frac{H}{\sin(\beta)},$$

$PQ = \frac{O_2P \times P_1Q_1}{O_2P_1}$ can be obtained according to $\frac{PQ}{P_1Q_1} = \frac{O_2P}{O_2P_1}$, thus obtaining the coordinate X = PQ in the horizontal direction. Therefore, the true coordinates of the point Q are (X, Y).

3. Detection of underground human body key-points

Conventional key-point detection algorithms mostly use a "top-down" approach, which first detects all persons in an image to be detected, and then detects key-points of each person separately, failing to realize high-speed detection for a large number of

Date Reçue/Date Received 2020-09-24

persons. In contrast, the present invention uses a "bottom-up" structure to first detect key-points of all persons in a video and then cluster key-points belonging to each individual by means of matching and optimization in a bipartite graph, thus achieving the objective of detecting human body key-points of each person in the video. In this way, the detection speed will not decrease with increase in the number of detected persons, realizing multi-person real-time detection of human body key-points. A key-point detection structure is shown in FIG. 4.

With a color RGB image as an input, deep features thereof are extracted via a 19-layer deep neural network ("VGG 19") to obtain a feature graph shown in FIG. 4. Afterwards, the feature graph is input to step 1 having two CNN branches. In this step, the first network branch predicts confidence maps $S^1=\rho^1(F)$ regarding a set of key-points, where ρ^1 indicates an inference procedure of the first network branch in step 1. The second network branch predicts a set of "part affinity fields" $\mathbf{L}^1= \emptyset^1(\mathbf{F})$, where $\emptyset^1$ indicates an inference procedure of the second network branch in step 1 and aims to cluster the predicted key-points on respective limbs of each individual, to obtain a set of complete information of human body key-points. The prediction results obtained by this step based on the two network branches are linked in series with the original feature graph, and then are together input to subsequent steps t as shown in FIG. 4, so as to obtain a more accurate prediction result. The subsequent inference steps t may be denoted by the following formulas:

$$\mathbf{S} = \rho^t(\mathbf{F},\mathbf{S}^{t-1},\mathbf{L}^{t-1}), \quad \forall t \geq 2$$

$$\mathbf{L} = \emptyset^t(\mathbf{F},\mathbf{S}^{t-1},\mathbf{L}^{t-1}), \quad \forall t \geq 2$$

where ρ^t and $\emptyset^t$ indicate inference procedures of the two CNN branches in step t respectively. According to FIG. 4, t in the subsequent steps t is ≥ 2.

In order to guide the network to iteratively predict the confidence maps and the "part affinity fields" of the key-points, a loss function L_2 (also referred to as a mean squared error) is applied after each step for each branch, which is used to measure an error between a predicted value and a true value. Herein, a spatial weighted value is used to solve the problem that some datasets are unable to mark the key-points of all persons. The loss functions for the CNN branches in step t may be denoted by the

Date Reçue/Date Received 2021-05-26

following formulas respectively:

$$f_S^t = \sum_{j=1}^{J} \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \left\| \mathbf{S}_j^t(\mathbf{p}) - \mathbf{S}_j^*(\mathbf{p}) \right\|_2^2$$

$$f_L^t = \sum_{c=1}^{C} \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \left\| \mathbf{L}_c^t(\mathbf{p}) - \mathbf{L}_c^*(\mathbf{p}) \right\|_2^2$$

where $\boldsymbol{S}_j^*$ indicates a true position of the key-point confidence map; $\boldsymbol{L}_c^*$ indicates a true position of the "part affinity field"; W is a Boolean value, which equals 0 when there is no mark in a dataset, or otherwise equals 1, and is used to avoid the detection network from punishing the true key-points without any mark.

A true reference for defining confidence levels of key-points at any position p in the picture is as follows:

$$\mathbf{S}_{j,k}^*(\mathbf{p}) = \exp\left(-\frac{\left\| \mathbf{p} - \mathbf{x}_{j,k} \right\|_2^2}{\sigma^2} \right)$$

where p is coordinates of any position in the picture to be detected, k indicates the kth person in the picture, $x_{j,k}$ indicates true coordinates of the jth key-point of the kth person in a marked training sample, and σ indicates a constant for controlling the degree of dispersion of a Gaussian graph regarding confidence points. FIG. 5 is a curve chart showing a relationship between key-point coordinates (k, j) and confidence levels.

A maximum value is taken from the foregoing chart, to obtain a confidence reference $\mathbf{S}_j(\mathbf{p}) = \max \boldsymbol{S}_{j,k}^*(\mathbf{p})$ of the jth key-point of the kth person.

4. Method for front-view and top-view projections of the key-point coordinates and the ROI

In the "bottom-up" approach for detection of key-points, coordinate information of all the key-points is output in the end. The above-mentioned monocular visual method can calculate x_w-axis and y_w-axis components in the world coordinate system that correspond to a particular pixel coordinate point in the video, which are enough for front-view and top-view projections. However, if it is unable to calculate a z-axis

Date Reçue/Date Received 2020-09-24

component of a human body key-point, projection of the key-point in the front-view direction cannot be implemented. In addition, because a target pixel in a monocular visual image does not contain depth information capable of reflecting a 3D relationship, transformation from the image coordinate system to the world coordinate system cannot be realized. To solve such a problem, the present invention simplifies a human body key-point model as follows: According to the already known x_w-axis and y_w-axis components of the key-points in the world coordinate system, a height component z_w is self-defined for each key-point based on the key-point model shown in FIG. 7, and the defined height component and the known x_w-axis and y_w-axis components are then combined into complete world coordinates (x_w, y_w, z_w) of the human body key-point. Front-view and top-view projections of a belt dangerous region ROI corresponding to the key-point coordinates are shown in FIG. 6.

To shorten a system operation time, the human body key-point model is simplified. FIG. 7(a) shows an original human body key-point model of system prediction, where there are 25 key-points in total. Some key-points in the original model are omitted and only key-points numbered 0, 2, 5, 4, 8, 7, 22 and 19 are kept, thus simplifying the original model to obtain a simplified model shown in FIG. 7(b).

In this model, the z_w-axis component of the point 0 is set to 1.6m; the z_w-axis components of the points 1, 2, and 5 are all set to 1.3m; the z_w-axis components of the points 4, 8, and 7 are set to 1m; and the z_w-axis components of the points 22 and 19 are set to 0m because they are on the same plane as the belt. A projection effect is shown in FIG. 8 where (a) is a top-view projection and (b) is a front-view projection.

For the simplified human body key-point model, if the minimum distances d_T and d_F from the key-points to the dangerous region ROI respectively in the top-view and front-view projections are both less than or equal to a safe distance threshold d, and h in the front view is less than the height (1.5m) of the belt dangerous region, the system evaluates a current action as an unsafe action and gives an early warning.

5. Method for identifying a dangerous action of an underground operator

The above-described method for evaluating an unsafe action based on the positional

Date Reçue/Date Received 2020-09-24

relationship fails to determine a specific type of the unsafe action, for example, fails to determine that the operator falls over beside the equipment or leans against or sit on the equipment. All these actions have a great safety hazard. Therefore, identification of a specific action of an underground conveyor belt operator is an issue in urgent need to be solved.

The present invention further identifies a specific type of an action in addition to evaluating a dangerous action based on the positional relationship; and sets different safe distance thresholds according to different levels of danger of the actions.

A simple DNN classifier is established, and used to classify actions according to the key-point information acquired in the above. The position information of the key-points in each picture is combined into a sample which corresponds to one action type. A large number of marked key-point - action samples are used to train the classifier, such that the classifier has the ability to identify human actions in each single picture. In addition, considering continuity of actions of the operator in a surveillance video, a probability judgment model for multiple successive pictures which are usually highly correlated is added based on the action identification for each single picture. Specifically, by using five successive pictures as a unit for judgment, results of classifying actions in the five pictures one by one are returned with the action classifier, and counting is performed for different classification results separately. Finally, a ratio of the counted number in each result to the total number in all the results is calculated, and the greatest ratio is determined as a classification result of the actions in the five pictures. A procedure of classifying the actions of underground operators by the classifier based on the DNN is shown in FIG. 9.

Unsafe actions to be identified include falling over, squatting down, and smoking. The three action types all have varying degrees of impacts on safety of the conveyor belt operator. Therefore, different safety coefficients are set respectively for the three actions, where $\gamma_1 = 2.0$ for falling over, $\gamma_2 = 1.5$ for squatting down, and $\gamma_3 = 1.3$ for smoking. Safe distance thresholds corresponding to the different actions may be obtained by calculation with the formula $d_i = \gamma_i * d$ ($i = 1, 2, 3$). By combining action identification and position evaluation, the operator is pre-warned about an unsafe

Date Reçue/Date Received 2020-09-24

distance according to different actions with different safety coefficients, thus realizing advanced warning about dangerous actions and greatly improving reliability of the safety early-warning system. FIGs. 10 to 12 are schematic diagrams of safety judgment for the three actions respectively.

When horizontal distances d_{Ti} and d_{Fi} between each of the three dangerous actions and the belt are less than respective safety thresholds d_i, and further a vertical height h_i from a key-point horizontally nearest to the belt to the horizontal plane is less than the height of the ROI on the belt, the system determines a current action as an unsafe behavior and gives an alarm.

FIG. 13 shows a specific implementation process of the present invention, where (a) is a detection diagram of a belt dangerous region and human body key-points, (b) is a top view of the human body key-points and the belt dangerous region, and (c) is a front view of the human body key-points and the belt dangerous region.

The foregoing embodiment merely describes the technical idea of the present invention, but is not intended to limit the protection scope of the present invention. Any modification made based on the technical solutions according to the technical idea provided by the present invention falls within the protection scope of the present invention.

Date Reçue/Date Received 2020-09-24

What is claimed is:

1. A safety monitoring and early-warning method for a man-machine interaction behavior of an operator of an underground conveyor belt, comprising the following steps:

(1) acquiring an underground real-time video by using a surveillance camera, wherein the underground real-time video comprises video frames;

(2) estimating a size of a portion of the underground conveyor belt shown in the video according to a camera calibration principle, and accordingly delineating a three-dimensional region of interest(ROI) as a belt dangerous region;

(3) detecting human body key-points of all persons appearing in the video, measuring a degree of correlation between the human body key-points by using part affinity fields, and clustering key-points belonging to each person by means of matching and optimization in a bipartite graph, thus achieving an objective of detecting humanbody key-points of each person in the video;

(4) determining x-axis and y-axis components of each of the detected human body key-points in a world coordinate system, self-defining a height component z for each of the detected human body key-points, and then combining the x-axis and y-axis components and the height component z into complete world coordinates of the human body key-point; and

(5) determining, according to a positional relationship between the belt dangerous region and the detected human body key-points of each person, whether the man-machine interaction behavior is safe, thus determining whether to give an early warning.

2. The safety monitoring and early-warning method according to claim 1, wherein in step (3), using each of the video frames in the video as an input, deep features thereof are extracted to obtain a feature graph F, and then the feature graph F is input to step 1 having a first and a second convolutional neural network

Date Reçue/Date Received 2021-05-26

(CNN) branches; wherein the first CNN branch predicts confidence maps expressed as $S^1=\rho^1(F)$ regarding a set of the detected human body key-points, wherein ρ^1 indicates an inference procedure of the first CNN branch in step 1; the second CNN branch predicts a set of part affinity fields expressed as $\mathbf{L}^1= \emptyset^1(\mathbf{F})$, wherein $\emptyset^1$ indicates an inference procedure of the second CNN branch in step 1 and aims to cluster predicted human body key-points on respective limbs of each person, to obtain a set of complete information of each person's human body key-points; and afterwards, prediction results obtained by step1 based on the first and the second CNN branches are linked in series with the feature graph F, , which are then together input to subsequent steps t, so as to obtain further prediction results, wherein the further prediction results from each of the subsequent steps t are denoted by the following formulas:

$$\mathbf{S}^t = \rho^t(\mathbf{F},\mathbf{S}^{t-1},\mathbf{L}^{t-1}), \quad \forall t \geq 2$$
$$\mathbf{L}^t = \emptyset^t(\mathbf{F},\mathbf{S}^{t-1},\mathbf{L}^{t-1}), \quad \forall t \geq 2$$

wherein t is ≥ 2, each of the subsequent steps t comprises two CNN branches to predict confidence maps expressed as S^t and part affinity fields expressed as L^t respectively, and ρ^t and $\emptyset^t$ indicate inference procedures of the two CNN branches in each of the subsequent steps t respectively.

3. The safety monitoring and early-warning method according to claim 2, wherein a loss function, identified as a mean squared error, is applied for the two CNN branches separately after each of the subsequent steps t, and the loss functions for the two CNN branches in each of the subsequent steps t are respectively as follows:

$$f_S^t = \sum_{j=1}^{J}\sum_{\mathbf{p}} \mathrm{W}(\mathbf{p}) \cdot \left\|\mathbf{S}_j^t(\mathbf{p}) - \mathbf{S}_j^*(\mathbf{p})\right\|_2^2$$
$$f_L^t = \sum_{c=1}^{C}\sum_{\mathbf{p}} \mathrm{W}(\mathbf{p}) \cdot \left\|\mathbf{L}_c^t(\mathbf{p}) - \mathbf{L}_c^*(\mathbf{p})\right\|_2^2$$

Date Reçue/Date Received 2021-05-26

wherein f^t_S and f^t_L are respectively the loss functions for the two CNN branches in each of the subsequent steps t; p is coordinates of a position p in a video frame to be detected; W(p) is a Boolean value, which equals 0 when there is no mark in a training dataset, or otherwise equals 1; $\mathbf{S}^t j(\mathbf{p})$ indicates a confidence map of a jth human body key-point of a kth person at the position p in the subsequent step t; $\mathbf{S}^*_j(\mathbf{p})$ indicates a true position of the confidence map; $\mathbf{L}^t c(\mathbf{p})$ indicates a part affinity field at the position p in the subsequent step t, and $\mathbf{L}^* c(\mathbf{p})$ indicates a true position of the part affinity field; a true reference for defining confidence levels of key-points at any position p in the video frame is as follows:

$$\mathbf{S}^*_{j,k}(\mathbf{p}) = \exp\left(-\frac{\left\|\mathbf{p} - \mathbf{x}_{j,k}\right\|_2^2}{\sigma^2}\right)$$

wherein $x_{j,k}$ indicates true coordinates of the jth human body key-point of the kth person in a marked training sample, and σ indicates a constant for controlling a degree of dispersion of a Gaussian graph regarding confidence points; and

a maximum value is taken, to obtain a confidence reference $\mathbf{S}_j(\mathbf{p}) = \max S^*_{j,}$ **(p)** of the jth human body key-point of the kth person.

4. The safety monitoring and early-warning method according to claim 1, wherein each person has 9 human body key-points in total, which are each person's nose, chest, right shoulder, right hand, left shoulder, left hand, hip bone, rightfoot, and left foot.

5. The safety monitoring and early-warning method according to claim 1, wherein in step (5), front-view and top-view projections of the belt dangerous region are determined based on the belt dangerous region obtained in step (2); for each person in the video, a minimum distance d_T from the human body key-points to the top-view projection of the belt dangerous region, a minimum distance d_F

Date Reçue/Date Received 2021-05-26

from the human body key-points to the front-view projection of the belt dangerous region, and the height h of the corresponding human body key-point are calculated; and if d_T and d_F are both less than or equal to a safe distance threshold d and h is less than the height of the front-view projection of the belt dangerous region, it is determined that a man-machine interaction behavior of the operator is unsafe and an alarm is given.

6. The safety monitoring and early-warning method according to claim 5, wherein in step (5), a Deep Neural Network (DNN) classifier is established and used to classify actions according to the detected information of the human body key-points; position information of the human body key-points in each of the video frames is combined into a sample which corresponds to one action type;a large number of marked human body key-point - action samples are used to train the classifier, such that the classifier identifies human actions in each of the video frames; and safe distance thresholds d_i corresponding to different actions are determined according to identification results of the classifier, wherein i in the safe distance thresholds d_i represents an ith action type.

7. The safety monitoring and early-warning method according to claim 6, wherein considering continuity of actions of the operator in the video, a probability judgment model for multiple successive video frames is added based on the action identification for each single video frame: by using M successive video frames as a unit for judgment, returning results of classifying actions in the M successive video frames one by one with the classifier, and performing counting for different classification results separately; and finally, calculating a ratio of the counted number in each of the classification results to a total number in all the classification results, and determining the greatest ratio as a classification result of the actions in the M successive video frames.

8. The safety monitoring and early-warning method according to claim 6, wherein there are three action classification results: falling over, squatting down,

Date Reçue/Date Received 2021-05-26

and smoking; different safety coefficients γ_i are assigned to the three action types respectively, and their respective safe distance thresholds are calculated as follows: $d_i = \gamma_i \bullet d$, wherein i = 1, 2, 3; and finally, it is determined, according to the safe distance threshold, whether man-machine interaction of the operator for a current action is safe.

Date Reçue/Date Received 2021-05-26

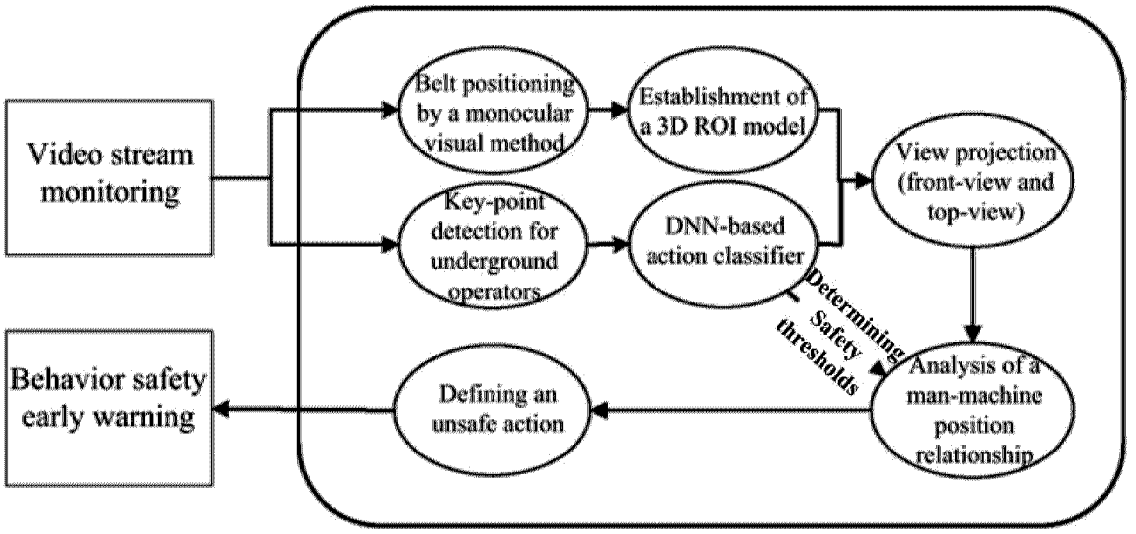


FIG. 1

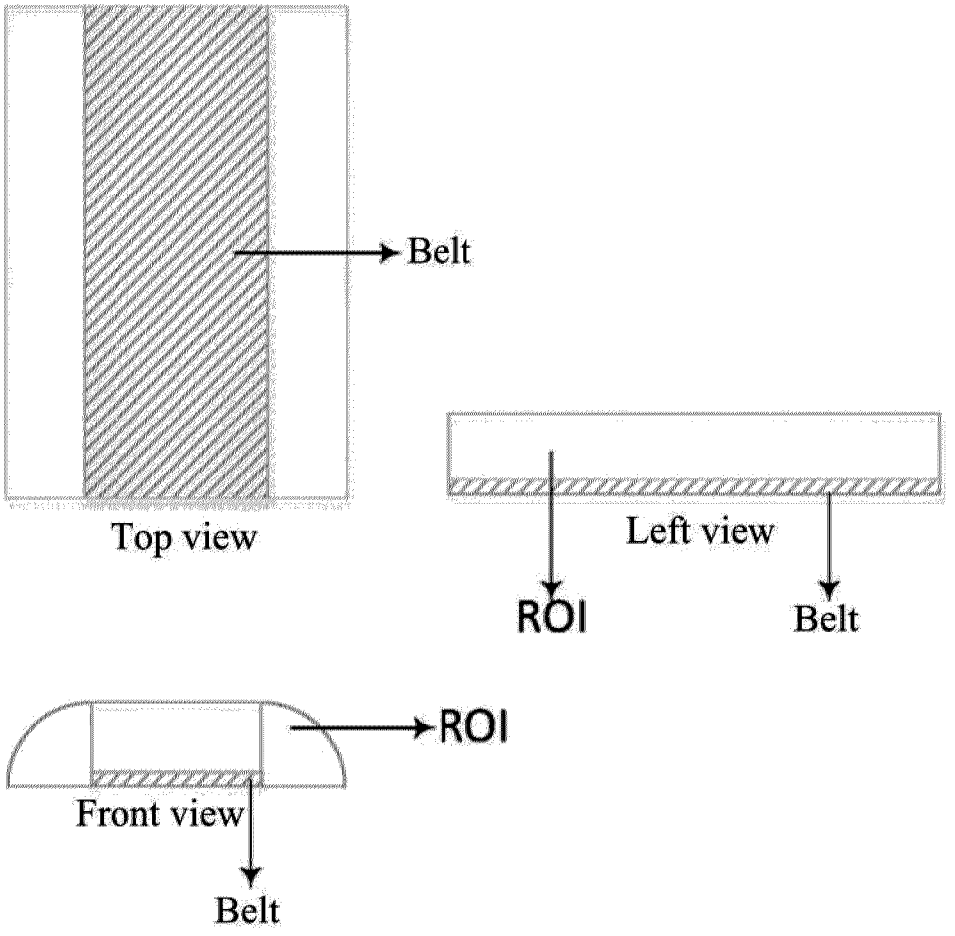


FIG. 2

Date Reçue/Date Received 2020-09-24

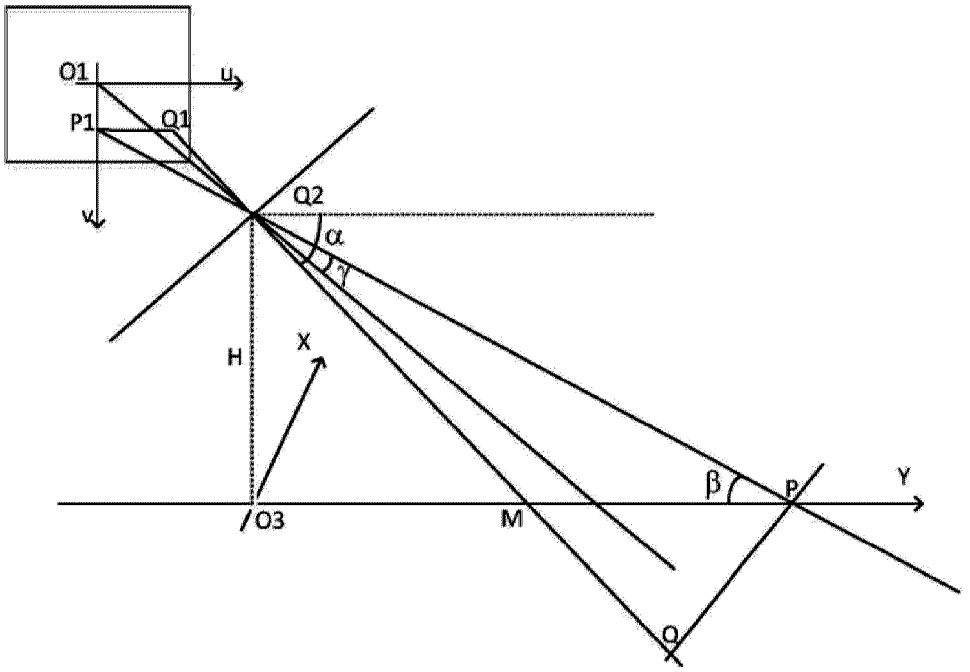


FIG. 3

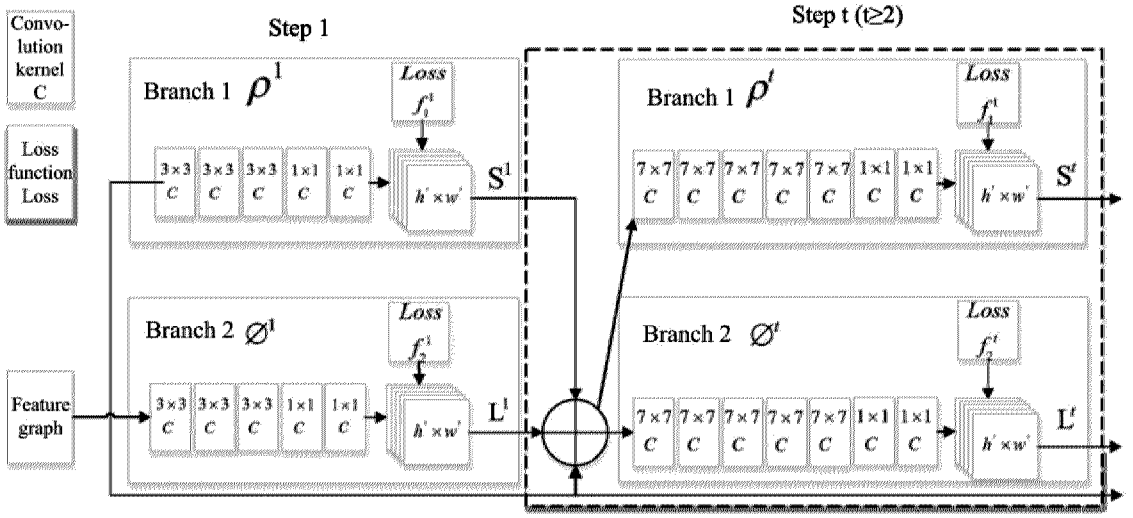


FIG. 4

Date Reçue/Date Received 2020-09-24

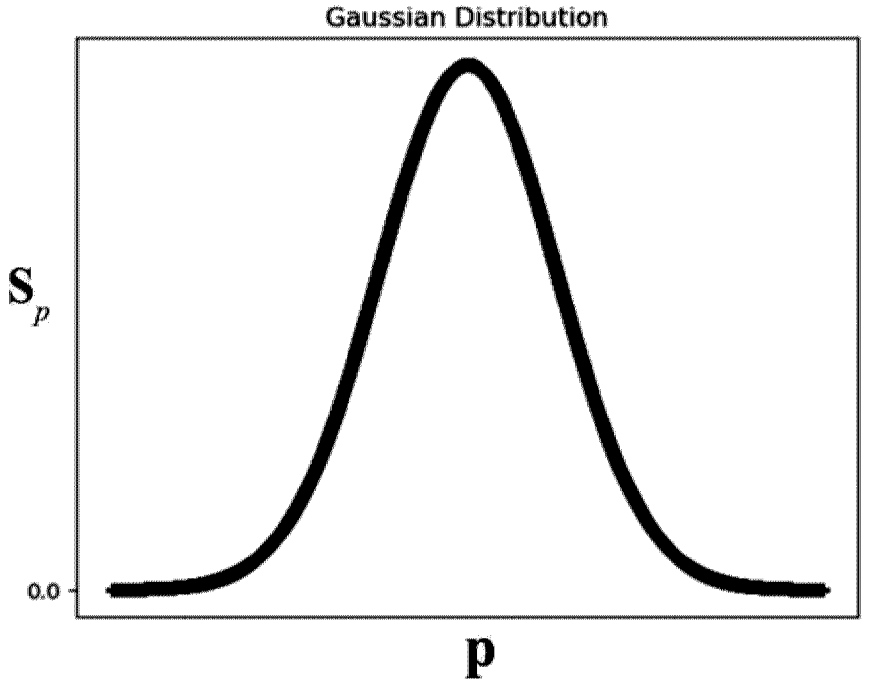


FIG. 5

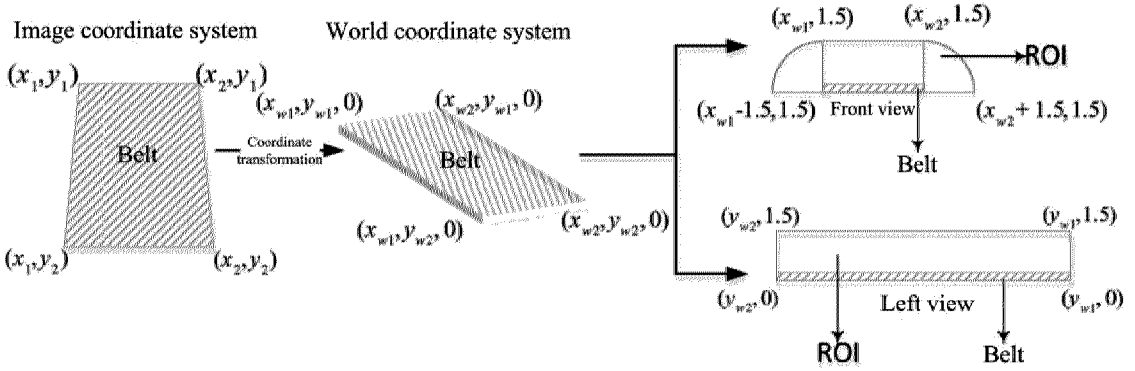
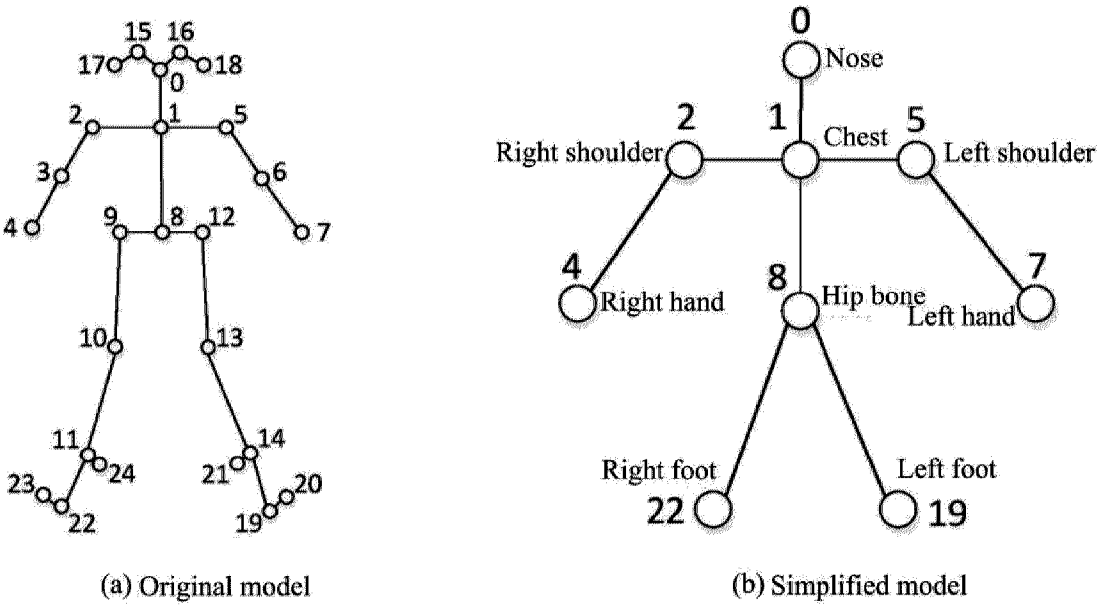


FIG. 6

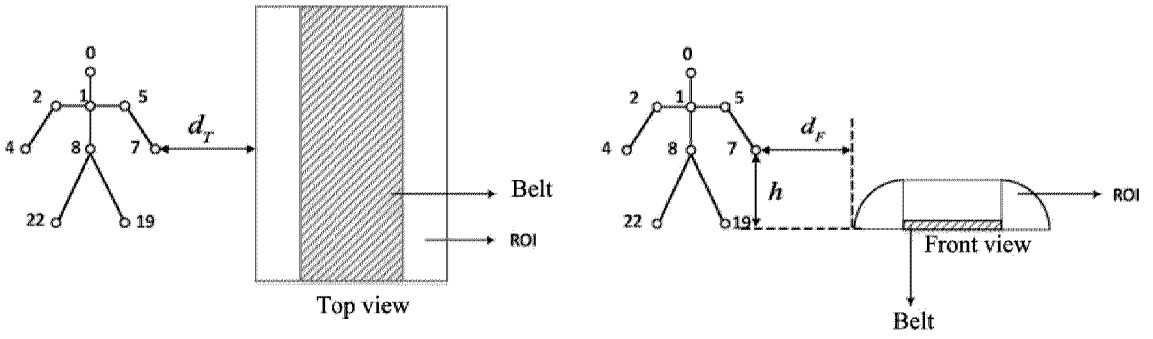
Date Reçue/Date Received 2020-09-24



(a) Original model

(b) Simplified model

FIG. 7



(a) Estimation of a top view

(b) Estimation of a front view

FIG. 8

Date Reçue/Date Received 2020-09-24

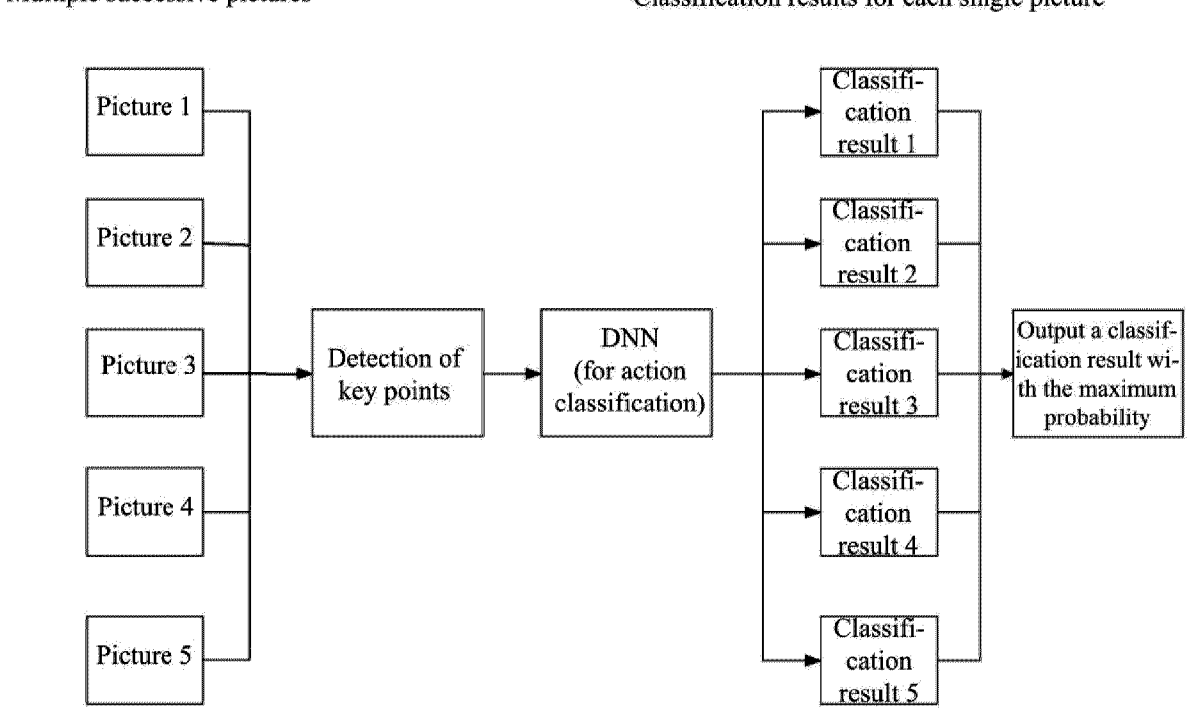


FIG. 9

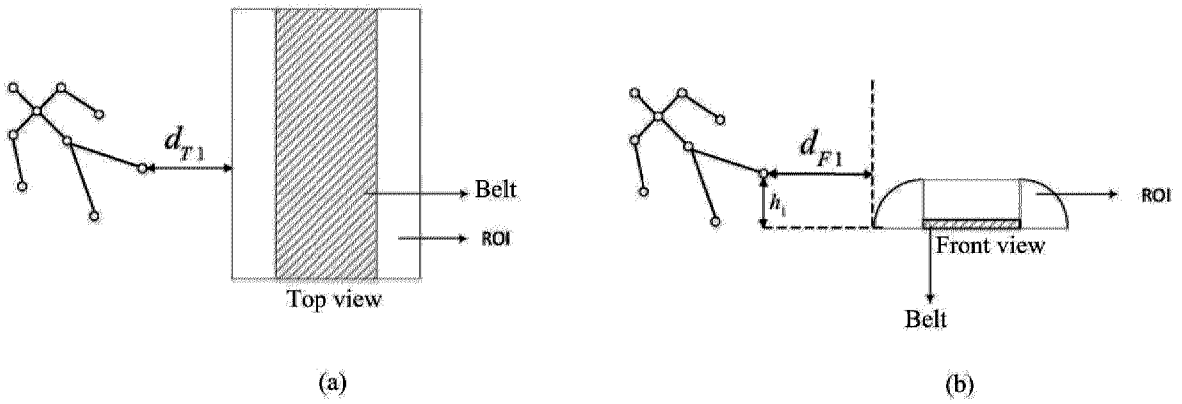


FIG. 10

Date Reçue/Date Received 2020-09-24

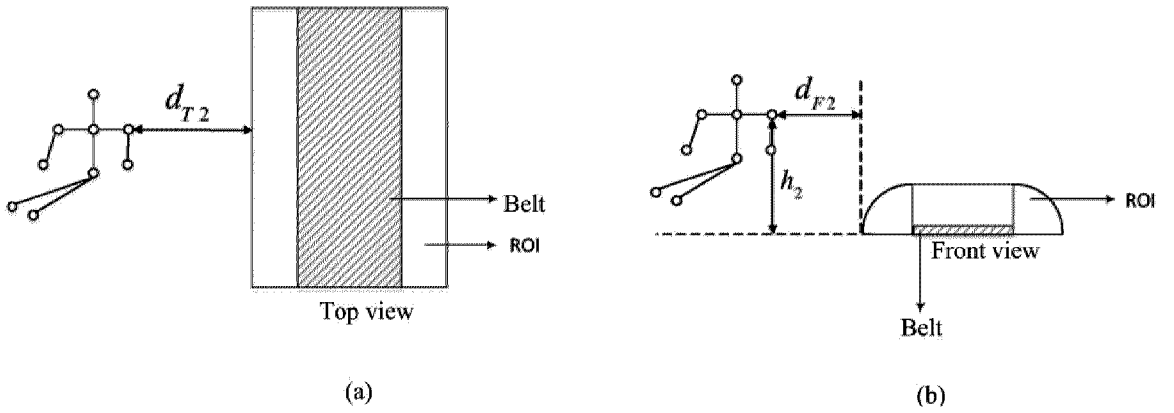


FIG. 11

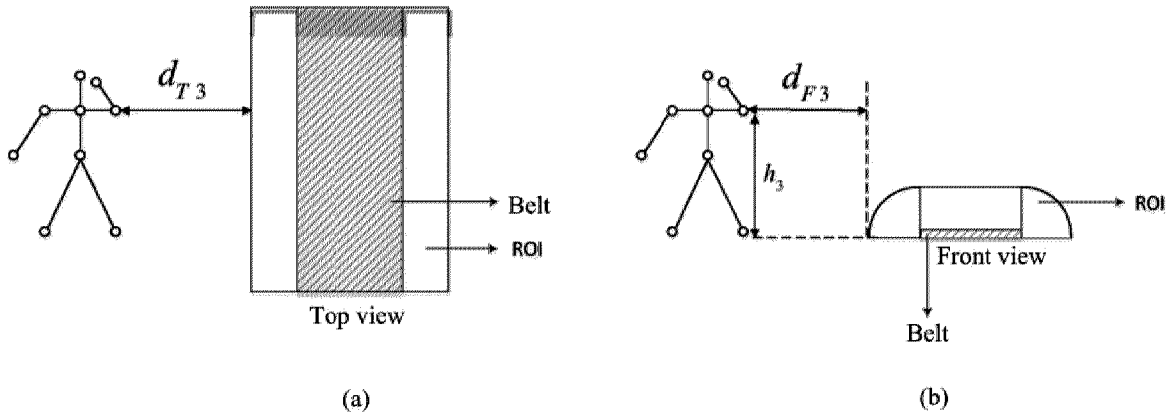
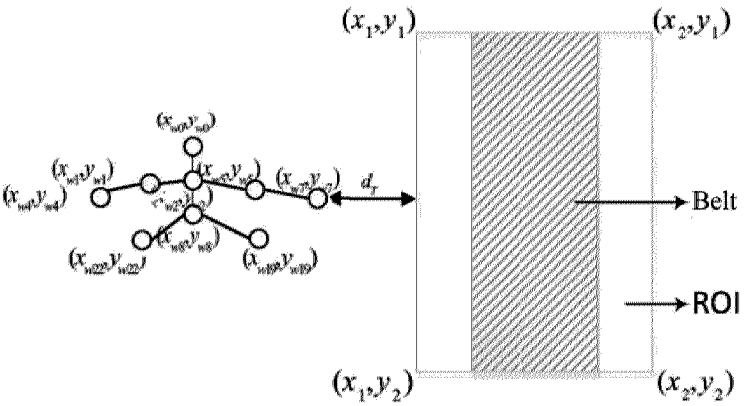
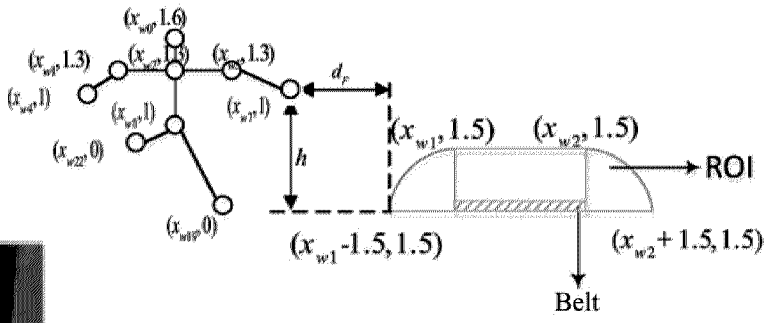


FIG. 12

Date Reçue/Date Received 2020-09-24



(b) Top view

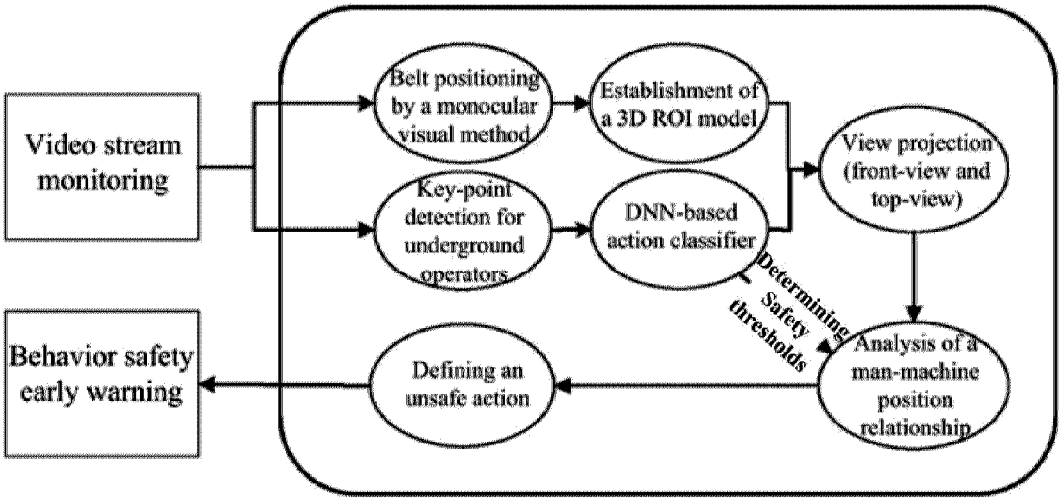


(c) Front view

(a) Detection of key-points of a conveyor belt operator and belt position

FIG. 13

Date Reçue/Date Received 2021-05-26


Video stream monitoring
Belt positioning by a monocular visual method
Establishment of a 3D ROI model
View projection (front-view and top-view)
Key-point detection for underground operators
DNN-based action classifier
Determining safety thresholds
Analysis of a man-machine position relationship
Defining an unsafe action
Behavior safety early warning