3D Reconstruction of Underground Tunnel Using Depth-camera-based Inspection Robot

Ningbo Jing,[1*] Xianmin Ma,[1] Wei Guo,[2] and Mei Wang[1]

[1]College of Electrical and Control Engineering, Xi'an University of Science and Technology,
Xi'an 710054, China
[2]College of Mechanical Engineering, Xi'an University of Science and Technology,
Xi'an 710054, China

Establishing a 3D model of an underground environment for an inspection robot has received significant attention and concern in recent years. RGB and depth images are obtained using a depth camera. The acquired RGB and depth maps are filtered to remove noise points using a Markov random field (MRF)-based filter. A novel deep neural network (DNN) architecture that implements the feature description is proposed. The feature points of a depth image are extracted to realize the precise matching between the RGB and depth images. Point clouds are obtained and registered into a single position using an improved iterative closest point algorithm. The experimental results show the effectiveness and practicability of the proposed method. An accurate 3D reconstruction of the object has been achieved with a dense point cloud.

1. Introduction

The unmanned production of underground coal mines is the most effective way to solve coal mine safety problems. The coal mining environment and the particularity of operations are the main factors restricting the development of underground robots. The underground structure of a coal mine is intricate with nonuniform illumination. The lack of an accurate map in underground mines is a serious threat to the safety of both public and mine workers. Realizing the autonomous navigation of an underground robot is inseparable from the 3D reconstruction of an underground tunnel environment. Establishing a 3D model for an underground environment has received significant attention and concern in recent years. Domestic and foreign scholars have researched this field.

Li and Zhan introduced intelligent and unmanned control technologies for underground mines. Tested unmanned devices were equipped with a wireless communication system, location and navigation systems, and a data acquisition system.[1] Jing *et al.* introduced the 3D reconstruction of an underground tunnel using a Kinect camera.[2] Troubleshooting and safety monitoring in an underground mine by manual operation often have security risks, so the trend of replacing manual operation with robots increases. The navigation system plays an

important role in mine rescue robots in underground mine disasters, Tian *et al.* proposed a new navigation method with diverse-sensor data fusion using an improved algorithm of a neural-network-extended Kalman filter.[3] Intelligent mining has revolutionized the coal industry. They have realized an unmanned operation along the work face using a remote monitoring video on a roadway and a device operated automatically and monitored using a remote-control center.[4] Yang *et al.* proposed a new close-range photographer control method for the roadway excavating face images of coal mines, which is greatly significant for the safe production and resource recycling of coal mines.[5] Zhang *et al.* proposed a framework for the classification and reconstruction of point cloud data for a large number of objects.[6] The reconstruction of 3D building models is still challenging in 3D city modeling. The process starts with the segmentation of point clouds of roofs and walls into planar groups. By generating related surfaces and using geometrical constraints with the symmetry considered, one can reconstruct a 3D building model.[7] Dai *et al.* proposed a real-time globally consistent 3D reconstruction using on-the-fly surface reintegration.[8] The Zhang[9] method was used for sensor calibration. Camplani *et al.*[10] presented a depth-color fusion strategy for the 3D modeling of indoor scenes with Kinect. Accurate depth and color models of the background elements are iteratively quilted in the scene. Kinect depth data is processed with an innovative adaptive joint-bilateral filter that efficiently combines depth and color by analyzing an edge-uncertainty map and the detected foreground regions.

Inspection Robotics delivers products with industrial-grade reliability and robustness that meet high quality standards. Inspection robots enable operators to plan outages more precisely and efficiently, thereby reducing downtime as well as boosting the safety of people and the environment. The 3D structure reconstruction of an underground environment is realized using a Kinect depth sensor. The rest of this paper is organized as follows. In the next section, the system hardware platform and calibration method of the camera are introduced. In Sect. 3, the RGB and depth image enhancement method under nonuniform illumination, Markov random field (MRF)-based filters, and a novel deep network architecture for feature point detection, orientation estimation, and feature description are introduced. In Sect. 4, experimental results are given. Finally, conclusions are given in Sect. 5.

2.   Sensor Calibration

Sensor calibration is a method of improving sensor performance by removing structural errors in sensor outputs. Sensors need to be calibrated in the systems they are being used for. Structural errors are differences between the expected and measured outputs of sensors, which occur consistently every time a new measurement is taken. Camera calibration is the process of estimating intrinsic and extrinsic parameters. Intrinsic parameters are the internal characteristics of the camera, such as focal length, skew, distortion, and image center. Extrinsic parameters describe the position and orientation of the camera in the environment.

A time-of-flight (ToF) camera sensor on a drone or the ground has many powerful uses. It is a range imaging camera system that resolves distance on the basis of the known speed of light, measuring the ToF of a light signal between the camera and the subject for each point of

the image. The artificial illumination may be provided by a laser or an LED. Laser-based ToF cameras are part of a broader class of scannerless light detection and ranging (LIDAR) systems, in which the entire scene is captured with each laser pulse, as opposed to point-by-point with a laser beam such as in scanning LIDAR systems. The simplest version of a ToF camera uses multiple light pulses or a single light pulse. The illumination is switched on for a very short time, the resulting light pulse illuminates the scene and is reflected by the objects in the field of view. The camera lens gathers the reflected light and images it onto the sensor or focal plane array. Depending on the distance, the incoming light experiences a delay. ToF camera products for civil applications began to emerge around 2000 as the semiconductor processes became fast enough for such devices. The systems cover ranges of 5 cm to 2 km. The distance resolution is about 1 cm. The spatial resolution of ToF cameras is generally lower than that of standard 2D video cameras, with most commercially available devices at 320×240 pixels or less. Compared with other 3D laser scanning methods for capturing 3D images, ToF cameras operate very rapidly, providing up to 160 images per second.

2.1 System hardware platform

In this experiment, we used a Kinect v2 camera on a coal mine robot. Its structure is shown in Fig. 1. It contains an IR laser emitter, an RGB camera, and an IR Camera. The IR laser emitter is used to emit infrared light, the IR camera to obtain depth images, and the RGB camera to obtain color images. Kinect for Windows SDK includes a driver, a rich Raw Sensor Streams API, a natural user interface, an installation file, and reference data. Kinect for Windows SDK makes it easy for programmers to use C++, C or Visual Basic with Microsoft Visual Studio tools. Kinect for Windows SDK provides raw sensing data streams that developers can access directly from distance sensors, color cameras, and four-cell microphone arrays. These data streams allow developers to develop applications based on the raw data stream of the Kinect sensor.

The main body of the experimental tracked robot is a deformable tracking system composed of four rotating arms, which can rotate independently, and a track (Fig. 2). A deformable crawler system can use a rotating arm for posture control and keep its advantageous capability

Fig. 1.   Kinect v2 camera.

Fig. 2.   (Color online) Robot experiment platform.

to cross obstacles. It is the core factor of movement with all the advantages of supermobility and easy control in any terrain. With this device, the mobile robot can change its posture without disrupting track synchronization. A control panel can be installed on the fuselage. To drive the robot, six DC motors, which have the advantages of using the rotating arm and the track to move installed. A total of 4 microprocessors (one host and three slave machines) have been installed on the control board of the entire system. The main engine controls the entire operation of the robot. To control the motor, the three slave machines measure two encoders and provide them to the main engine. The control board is connected to a range sensor and a tilt sensor mounted on the body of the mobile robot, to detect obstacles and the state of the fuselage. In addition, measuring the value of the encoder installed on the motor can control the speed and posture of the robot.

2.2　Calibration of internal parameters

In this section, we describe in detail the method of matching the RGB and infrared images collected by the Kinect camera. To obtain better corners, the red thin iron plate shown in Fig. 3 is adopted. The white part is a hole. When the depth image is perpendicular to the iron plate, the pixels of the image are different when the distance between the holes is different. In this manner, better corner information can be obtained. A sample image consisting of an infrared image and an RGB image is shown in Fig. 3. The obtained depth image is shown in Fig. 4(b). There are small black stripes at the edge of the hole, and no good corner can be obtained. To
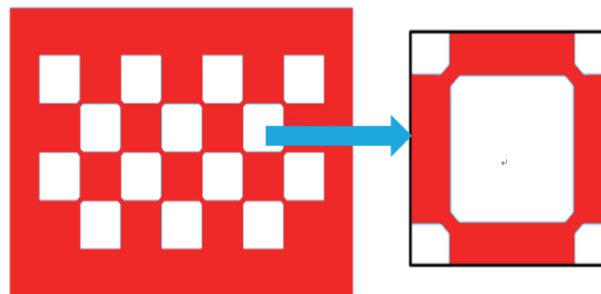

Fig. 3.　(Color online) Plate for correcting depth and RGB images.
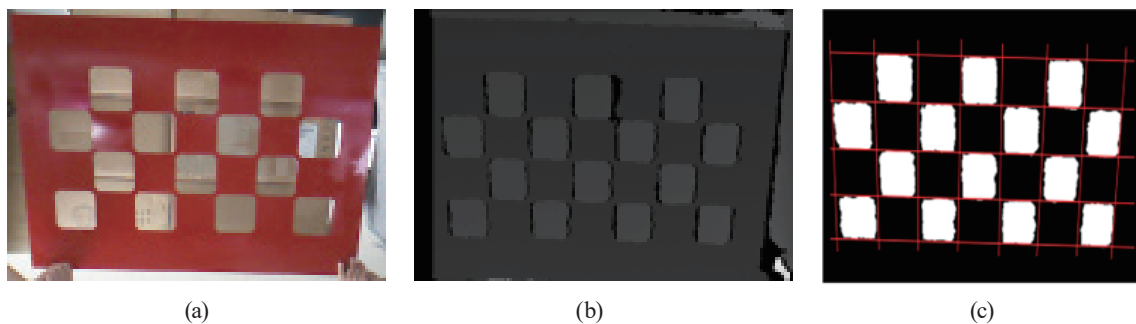

(a)　　　　　　　　　　　　　　　(b)　　　　　　　　　　　　　　　(c)

Fig. 4.　(Color online) (a) RGB image, (b) depth image, and (c) intersection lines of the experimental setup.

obtain better corner information, the following steps are adopted. As shown in Fig. 4(c), the corners of the red line on the two-value image are the characteristic corners on the graph.

Step 1: Smooth the image and adopt median filtering.
Step 2: Obtain the two-value image of an infrared image.
Step 3: On the binary image, manually select corners; each row clicks eight corners and each column clicks five corners, producing a total of 40 corners.
Step 4: The 40 corners manually selected on the basis of the least-squares method are used to obtain 13 straight lines, namely, five horizontal and eight vertical lines. With each straight line as the center, six pixel strip areas are extended outward to fit the straight line. When the overall error of the curve is less than the threshold, the iteration terminates. The camera matrix maps the 3D world scene into the image plane, the relationship is given by Eq. (1), where $[R\ T]$, called the extrinsic parameters, and K, called the camera intrinsic matrix, is given by Eq. (2).

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \tag{1}$$

$$\text{where} = \begin{bmatrix} \alpha_x & \gamma & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{2}$$

2.3 Calibration of extrinsic parameters

To solve the problems of the general corner detection algorithm in the check board, an improved corner detection algorithm based on the binary image binary corner in Zhang Zhengyou's calibration method is proposed.

Step 1: Binary image

In the raw data collected using the camera system, each pixel is RGB color data. The original 24 bit information is expressed as 1 bit, which is beneficial for the real-time processing of an image by the computer. The two-value image can highlight the contour of the target of interest. The gray value of each pixel on the gray image is set to 255 or 0, and the threshold is different in each pixel. The weighted average value of the 21×21 area around the pixel is calculated.

Step 2: Erosion and dilation

In a binary morphology, an image is viewed as a subset of a Euclidean space. The basic idea of the binary morphology is to probe an image with a simple, predefined shape, drawing conclusions on how this shape fits or misses the shapes in the image. This simple "probe" is called the structuring element and is itself a binary image. Erosion and dilation are two

fundamental operations for processing images on the basis of shape in morphological image processing from which all other morphological operations are based. Erosion was originally defined for binary images, later being extended to grayscale images and subsequently to complete lattices. Certain forms of structural elements are used to measure and extract the corresponding shapes in the image. Set *A* is used to represent the input image and Set *B* to represent the structural element. The result of an expansion by Set *B* is the union of all points of image *A* relative to structural element *B* after translation *B*. Expansion can fill holes smaller than the structural elements in the image and the small holes at the edge of the image. It has an external filtering effect on the image. Structure element *B* is used to expand the binary image to enlarge the white area and separate the black blocks. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels from object boundaries. The number of pixels added or removed from the objects in an image depends on the size and shape of the structure element used to process the image. In the morphological dilation and erosion operations, the state of any given pixel in the output image is determined by applying a rule to the corresponding pixel and its neighbors in the input image. The formula is as follows.

$$A \ominus B = \{z \in E \mid B_z \subseteq A\},\qquad(3)$$

where B_z is the translation of *B* by the vector *z*.

Step 3: Extraction contour and classification

If one point in the binary image is black and all eight adjacent points are black, the point is an inner part, and all inner points are hollowed out to get the outline of the image. After these steps, several quadrilateral outlines are screened out. Euclidean distance is calculated for all the vertices of the contour. If the distance between vertices is less than 2, the corresponding contour is an adjacent contour, and the count of the adjacent contours of this contour is added by 1. The common point in the target class is the corner.

Step 4: Subpixel corner coordinate extraction

The checkerboard corners detected by the above methods are not single pixel corner points, but their detection accuracy is low.

$$f(x,y) \approx a_0 + a_1 x + a_2 y + a_3 xy\qquad(4)$$

Here, the coefficients are determined by solving the linear system.

$$\begin{bmatrix} 1 & x_1 & y_1 & x_1 y_1 \\ 1 & x_1 & y_2 & x_1 y_2 \\ 1 & x_2 & y_1 & x_2 y_1 \\ 1 & x_2 & y_2 & x_2 y_2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} f(Q_{11}) \\ f(Q_{12}) \\ f(Q_{21}) \\ f(Q_{22}) \end{bmatrix}\qquad(5)$$

When the camera is precisely calibrated, the corner must be positioned to the subpixel level. Subpixel accuracy indicates the subdivision between two adjacent pixels. The two-degree polynomial interpolation algorithm is used to calculate the subpixel coordinates of the corners.

3.   Algorithm

Owing to the difference in the level of underground illumination, the images captured by the camera in different locations would either become very dim or bright depending on point light sources. Owing to the illumination, information extraction and decision making on the captured image would become difficult. Therefore, the enhancement of images before 3D reconstruction analysis becomes indispensable. Because the illumination environment in coal mines is not uniform, sharpening the captured image is the premise to improve the accuracy of 3D reconstruction. Owing to the difference in image processing space, general filter methods can be divided into frequency and spatial domains. In frequency domain enhancement, images are taken as two-dimensional signals, then signals are enhanced by domain transformation. The median filter method and partial averaging are two representative methods of frequency domain enhancement. In this paper, we introduce the use of a MRF-based filter algorithm to remove noise points of images.

3.1   Improved image enhancement method

Image enhancement technology plays a vital role in the 3D reconstruction of underground tunnels for inspection robots. The main purpose of image enhancement is to provide better results, as 3D reconstruction always prefers high-quality pictures to obtain the desired results. Image enhancement techniques are based on the following two broad categories: the spatial domain method, which operates directly on pixels, and the frequency domain method, which operates on the Fourier transform of an image. The spatial domain method concentrates on histogram equalization in an image enhancement field. The histogram equalization method is generally used to increase the contrast of many images, especially when the important data of those images are represented by close contrast values. Through this adjustment, the intensity can be better distributed across the histogram. Histogram equalization effectively spreads out the most frequent intensity values. On the other hand, wavelet transform is based on 2D discrete wavelet transform (2DDWT) (Fig. 5). A discrete wavelet transform (DWT) is another wavelet transform by which wavelets are discretely sampled for numerical and functional analyses. The advantage of DWT is that it captures both frequency and time information. The DWT decomposes signals into subbands with smaller bandwidths and lower
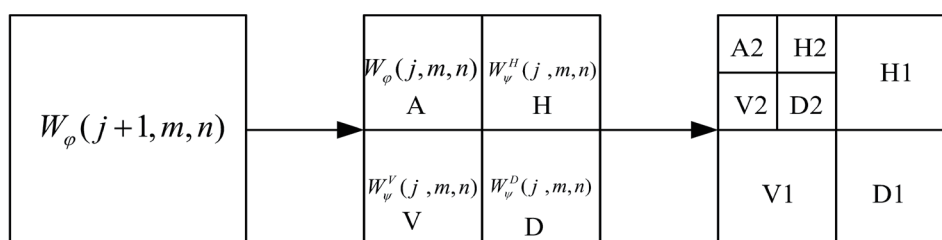

Fig. 5.   2D wavelet decomposition diagram.

sample rates, namely, low–low (LL), low–high (LH), high–low (HL), and high–high (HH). It is computationally impossible to analyze a signal using all wavelet coefficients, so one may wonder if it is sufficient to pick a discrete subset of the upper half-plane to be able to reconstruct a signal from the corresponding wavelet coefficients.

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a^m}} \psi\left(\frac{t-nb}{a^m}\right) \tag{6}$$

$$x(t) = \sum_{m\in\mathbb{Z}}1\sum_{n\in\mathbb{Z}}1\langle x,\psi_{m,n}\rangle \cdot \psi_{m,n}(t) \tag{7}$$

$$WT_{\psi}\{x\}(a,b) = \langle x,\psi_{a,b}\rangle = \int_{\mathbb{R}} x(t)\psi_{a,b}(t)dt \tag{8}$$

For n levels of decomposition, the wavelet packet decomposition (WPD) produces different $2n$ sets of coefficients (or nodes) as opposed to $(3n + 1)$ sets for the DWT. However, owing to the down sampling process, the overall number of coefficients is still the same and there is no redundancy. From the viewpoint of compression, the standard wavelet transform may not produce the best result, since it is limited to wavelet bases that increase by a power of two towards low frequencies. It is possible that another combination of bases produces a more desirable representation for a particular signal.

Image enhancement based on wavelet has greater advantages than conventional methods. Wavelet variation has multiresolution analysis characteristics. Images would have better clarity in details, highlights in subtle details, and a higher sense of depth. The wavelet method can also enhance the rough sketch of the original image. The functional block diagram of the enhancement algorithm can be seen in Fig. 6.

The information in the enhanced image shows a distorted process. Its purpose is to improve the wide dynamic problem of the images collected in coal mines. It meets the needs of not only the overexposed part but also the underexposed part caused by the low illumination. Finally, a balance state that meets the needs of 3D reconstruction is achieved. The image after improved filter enhancement can be seen in the picture, in which the overall brightness is improved without the overenhancement of brightness and the image effect is also greatly improved.

3.2　MRF-based filter

In the domain of artificial intelligence, a MRF is used to model various low- to mid-level tasks in image processing and computer vision. A Markov network or an undirected graphical model is a set of random variables with a Markov property described by an undirected graph. In other words, a random field is said to be a MRF if it satisfies Markov properties. A Markov network or an MRF is similar to a Bayesian network in its representation of dependences; their difference is that a Bayesian network is directed and acyclic, whereas a Markov network is undirected and may be cyclic. Thus, a Markov network can represent certain dependences that

```
                          ┌──────────────┐
                          │  Input image │
                          └──────────────┘
                                 │
                          ┌──────────────────────────┐
                          │ The decomposition coefficients are │
                          │ obtained by using two-dimensional  │
                          │ wavelet decomposition              │
                          └──────────────────────────┘
                                 │
                      ◇ To determine whether ◇
               N      ◇ the decomposition coefficient ◇
                      ◇ is greater than the threshold ◇
                                 │ Y
          ┌──────────────┐   ┌──────────────────┐
          │ Average threshold │   │ Local average energy │
          └──────────────┘   └──────────────────┘
                  │                  │
                  ┌──────────────────┐
                  │ Two-dimensional  │
                  │ wavelet reconstruction │
                  └──────────────────┘
                          │
                   ┌──────────────┐
                   │ Output image │
                   └──────────────┘
```
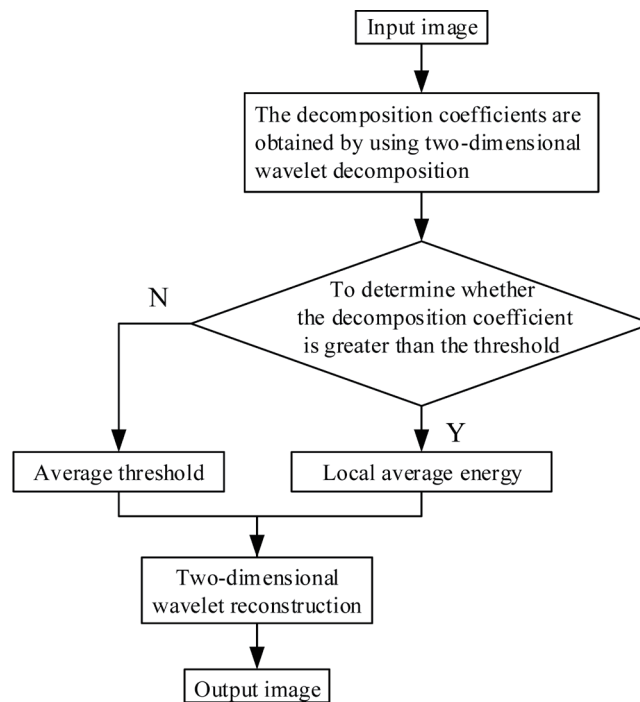
Fig. 6.   Block diagram of wavelet enhancement algorithm.

a Bayesian network cannot such as cyclic dependences; however, it cannot represent certain dependences that a Bayesian network can such as induced dependences.

In the probability theory, two events are independent, statistically independent, or stochastically independent if the occurrence of one does not affect the probability of occurrence of the other. Similarly, two random variables are independent if the realization of one does not affect the probability distribution of the other. Two random events A and B are conditionally independent given the third event C precisely if the occurrences of A and B are independent events in their conditional probability distribution given the third event C. In other words, A and B are conditionally independent given the third event C if and only if, given knowledge that C occurs, the knowledge of whether A occurs provides no information on the likelihood of B occurring, and the knowledge of whether B occurs provides no information on the likelihood of A occurring.

$$(A \perp\!\!\!\perp B)\,|\,C \Leftrightarrow Pr(A \cap B\,|\,C) = Pr(A\,|\,C)\,Pr(B\,|\,C) \tag{9}$$

The underlying graph of a MRF may be finite or infinite. Images are dissected into an assembly of nodes that may correspond to pixels or agglomerations of pixels. Hidden variables associated with the nodes are introduced into a model designed to "explain" the values (colors) of all pixels. A joint probabilistic model is built over pixel values and hidden variables. The

direct statistical dependences between hidden variables are expressed by explicitly grouping hidden variables; the obtained groups are often pairs depicted as edges in a graph. The concept of a hidden MRF model is derived from hidden Markov models. In statistics, a hidden MRF is a generalization of a hidden Markov model. Instead of having an underlying Markov chain, hidden MRFs have an underlying MRF. The main difference with a hidden Markov model is that the neighborhood is not defined in one dimension but within a network. In the vast majority of related literature, the number of possible latent states is considered a user-defined constant (Fig. 7).

When the joint probability density of random variables is strictly positive, it is also referred to as a Gibbs random field, because, according to the Hammersley–Clifford theorem, it can then be represented by a Gibbs measure for an appropriate energy function. The prototypical MRF is the Ising model; indeed, the MRF was introduced as the general setting for the Ising model. MRFs find application in various fields, ranging from computer graphics to computer vision and machine learning. MRFs are used in image processing to generate textures as they can be used to generate flexible and stochastic image models. In image modeling, the task is to find a suitable intensity distribution of a given image, where suitability depends on the type of task and MRFs are flexible enough to be used for image and texture syntheses, image compression and restoration, image segmentation, surface reconstruction, image registration, texture synthesis, super-resolution, stereomatching, and information retrieval. MRFs can be used to solve various computer vision problems, which can be energy minimization problems or problems where different regions should be distinguished using a set of discriminating features, within a MRF framework, to predict the category of the region. MRFs were a generalization over the Ising model and have, since then, been used widely in combinatorial optimizations and networks.

An MRF is a probability distribution over variables x defined by an undirected graph G in which nodes correspond to variables x.

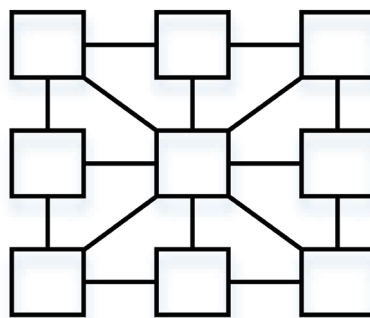$$P(X = x) = \frac{1}{Z} \exp\left(\sum_{k} w_k^{\mathrm{T}} f_k\left(x_{\{k\}}\right) \right) \tag{10}$$

Fig. 7.   Eight connected grids of image pixel Markov models in vision.

$$w_k^{\mathrm{T}} f_k\left(x_{\{k\}}\right) = \sum_{i=1}^{N_k} w_{k,i} \cdot f_{k,i}\left(x_{\{k\}}\right) \tag{11}$$

$$Z = \sum_{x \in \mathcal{X}} \exp\left(\sum_k w_k^{\mathrm{T}} f_k\left(x_{\{k\}}\right)\right) \tag{12}$$

The graphs corresponding to such MRF problems are predominantly gridlike, but may also be irregular. Graph connectivity is interpreted in terms of probabilistic conditional dependence. The importance of the partition function Z is that many concepts from statistical mechanics, such as entropy, are directly generalized in the case of Markov networks, and an intuitive understanding can thereby be gained. In addition, the partition function allows variational methods to solve the problems. One can attach a driving force to one or more of the random variables and explore the reaction of the network in response to this perturbation.

3.3   Deep neural network (DNN) architecture

Local features play a key role in 3D reconstruction. Finding and matching them across images have been the subject of vast amounts of research. Until recently, the best techniques relied on carefully hand-crafted features. Over the past few years, as in many areas of computer vision, methods based on machine learning, more specifically deep learning, have started to outperform conventional methods. Feature matching for static reconstruction can be learned using deep convolutional neural networks. Machine-learning-based approaches, especially in the form of DNNs, are a very promising avenue for tackling the many challenges in 3D reconstruction. Matching local geometric features on real-world depth images is a challenging task owing to the noisy, low-resolution, and incomplete nature of 3D scan data. The general expression of a convolution is

$$g(x,y) = (\omega * f)(x,y) = \sum_{s=-a}^{a} 1 \sum_{t=-b}^{b} w(s,t) f(x-s, y-t), \tag{13}$$

where g is the filtered image, f is the original image, and ω is the filter kernel. Every element of the filter kernel is considered as given by $-a \le s \le a$ and $-b \le t \le b$.

A DNN is an artificial neural network with multiple layers between the input and output layers. The DNN finds the correct mathematical manipulation to turn the input into the output, whether it be a linear or nonlinear relationship. Deep learning (deep structured or hierarchical learning) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semisupervised or unsupervised. Deep learning architectures such as deep neural, deep belief, and recurrent neural networks have been applied to computer vision. They have produced results comparable to and in some cases superior to those obtained by human experts. Deep

learning models are vaguely inspired by information processing and communication patterns in biological nervous systems yet have various differences from the structural and functional properties of biological brains (especially human brains), which make them incompatible with neuroscience evidence.

A biological neural network is composed of a group or groups of chemically connected or functionally associated neurons. A single neuron may be connected to many other neurons and the total number of neurons and connections in a network may be extensive. Connections, called synapses, are usually formed from axons to dendrites, although dendrodendritic synapses and other connections are possible. Apart from electrical signaling, there are other forms of signaling that arise from neurotransmitter diffusion. Artificial intelligence, cognitive modeling, and neural networks are information processing paradigms inspired by the way biological neural systems process data.

A convolutional neural network (CNN or ConvNet) is a class of DNNs that are most commonly applied to the analysis of visual imagery. CNNs use various multilayer perceptrons designed to require minimal preprocessing. They are also known as shift or space invariant artificial neural networks on the basis of their shared-weight architecture and translation invariance characteristics. Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

Neural networks are trained using stochastic gradient descent and require a loss function when designing and configuring the model. In classical statistics, sum-minimization problems arise in least-squares and maximum-likelihood estimations (for independent observations). The general class of estimators that arise as minimizers of sums is composed of M-estimators. However, in statistics, it has been long recognized that requiring even local minimization is too restrictive for some problems of maximum-likelihood estimation. Adam, a method for an efficient stochastic optimization, only requires first-order gradients with little memory requirement. This method computes individual adaptive learning rates for different parameters from estimates of the first and second moments of the gradients; the name Adam is derived from the adaptive moment estimation. The adaptive moment estimation is an update to the RMS prop optimizer. In this optimization algorithm, running averages of both the gradients and the second moments of the gradients are used. From the given parameters $w(t)$ and a loss function $L(t)$, where t indexes the current training iteration, Adam's parameter update is given by

$$m_w^{(t+1)} \leftarrow \beta_1 m_w^{(t)} + \left(1 - \beta_1\right) \nabla_w L^{(t)}, \tag{14}$$

$$v_w^{(t+1)} \leftarrow \beta_2 v_w^{(t)} + \left(1 - \beta_2\right)(\nabla_w L^{(t)})^2, \tag{15}$$

$$\hat{m}_w = \frac{m_w^{(t+1)}}{1-(\beta_1)^{t+1}},$$

(16)

$$\hat{v}_w = \frac{v_w^{(t+1)}}{1-(\beta_2)^{t+1}},$$

(17)

$$w^{(t+1)} \leftarrow w^{(t)} - \eta \frac{\hat{m}_w}{\sqrt{\hat{v}_w} + \varepsilon}.$$

(18)

Image features are defined in terms of local neighborhood operations applied to an image, a procedure commonly referred to as feature extraction, by which one can distinguish between feature detection approaches that produce local decisions, whether there is a feature of a given type at a given image point or not, and those that produce nonbinary data as a result. The distinction becomes relevant when the resulting detected features are relatively sparse. Although local decisions are made, the output from a feature detection step does not need to be a binary image. The result is often represented in terms of sets of (connected or unconnected) coordinates of the image points where features have been detected, sometimes with subpixel accuracy. By training correspondences from multiple existing RGB-D reconstruction datasets, each with its own properties of sensor noise, occlusion patterns, and various geometric structures and camera viewpoints, we can optimize the deep network to generalize and robustly match local geometries in real-world partial 3D data.

4.   Experiment and Result Analysis

The experiment is based on window 8 Intel (R) Core (TM) i5-4590 CPU @ 3.30 GHz RAM 4 64 bit OS. The resolution of RGB images obtained using the Kinect camera is 1080 × 1920 and the size of the depth image is 424 × 512. Firstly, the internal and external parameters of the camera are used to eliminate image distortion (Table 1). It can be seen that the enhancement method has many advantages over the conventional method in image enhancement. Some

Table 1
Internal and external parameters of camera.

No.	Parameters	Value	Reference
1	Focal length of X	364.785	
2	Focal length of Y	365.491	
3	Principal point X	257.812	Camera internal parameters
4	Principal point Y	207.425	
5	Radial distortion fourth order	0.271151	
6	Radial distortion second order	−0.0410451	
7	Radial distortion sixth order	−0.0874618	
8	Rotation matrix R	0.99999, 0.00108, 0.00398, 0.00181, 0.09998, 0.01711, 0.00501, 0.01601, 0.99991	Camera external parameters
9	Translation matrix T	0.02504, 0.00179, 0.00229	

specific indicators are mainly used for the objective evaluation of images. Error statistical results of the enhanced and original images are used to evaluate the quality of the enhanced image. In image processing, the mean value of image pixels reflects the brightness level of images. According to the representation method of the image gray level, the lower the image pixel mean value, the darker the image color and the higher the pixel mean value, the lighter the image color.

The mean value increases without increasing the noise signal. The image comentropy in the enhancement method is also higher than that in the conventional method. The enhanced image contains more information. The method effectively solved the problem of dimness and low contrast ratio of images in coal mines due to nonuniform illumination.

The experiment was completed in the advance support area of the underground tunnel using a single hydraulic prop in the main experimental center of Xi'an University of Science and Technology. The following results are shown in Fig. 8: (a) the RGB image and filtered RGB, (b) the depth image + the filtered depth image, (c) the 3D structure of the reconstruction map, and (d) the aerial view of the 3D reconstruction map.

It can be seen from Fig. 8 that the method effectively solved the problem of dimness and low contrast ratio of images in coal mines due to nonuniform illumination, and an accurate 3D structure map is obtained.
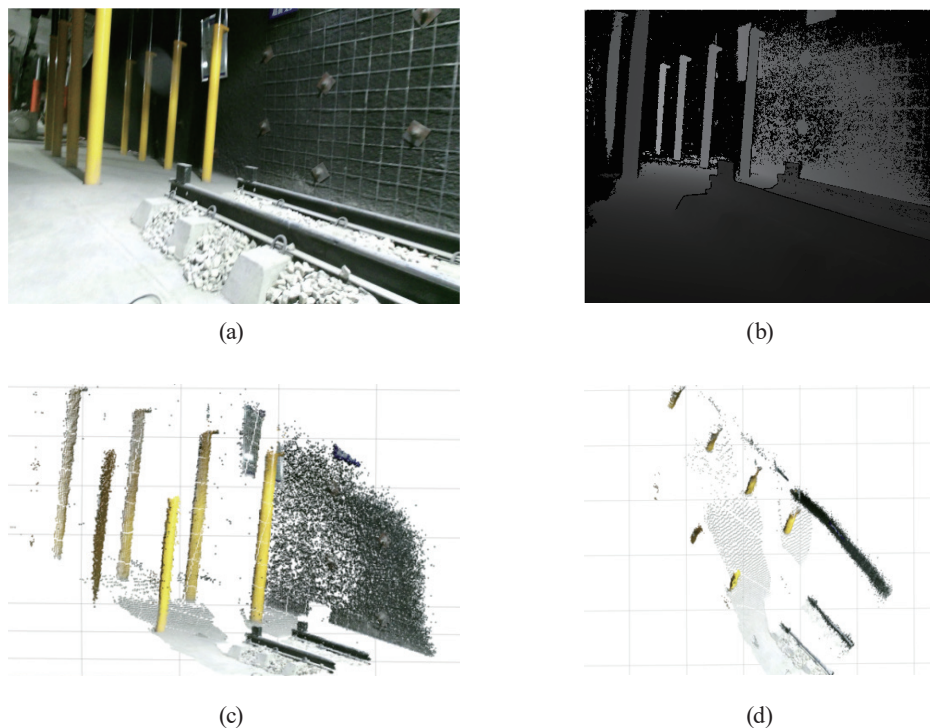

(a)　　　　　　　　　　　　　　　　　(b)

(c)　　　　　　　　　　　　　　　　　(d)

Fig. 8. (Color online) (a) RGB image, (b) depth image, (c) 3D reconstruction map, and (d) aerial view of 3D reconstruction map.

5. Conclusions

By focusing on images in coal mines and using the image enhancement algorithm, we realized the 3D structure reconstruction of an underground tunnel using the RGB and depth images obtained by a Kinect sensor. The feature corners of the depth image are adaptively extracted by an iterative method, and the camera image is corrected using the internal parameters of the camera. Then, the RGB and infrared images are enhanced and image matching and registration are realized by using a point cloud. The experimental results show the effectiveness and the feasibility of the proposed method. Further work on the one hand to improve the accuracy of the underground 3D reconstruction model is necessary; on the other hand, to realize an underground complex space robot, 3D path planning and autonomous navigation should be performed.

Acknowledgments

References

1 J. Li and K. Zhan: Engineering **4** (2018) 381. https://doi.org/10.1016/j.eng.2018.05.013
2 N. Jing, X. Ma, and W. Guo: Int. Symp. Computer, Consumer and Control (2018). https://doi.org/10.1109/is3c.2018.00077
3 Z. Tian, L. Zhang, and W. Chen: Comput. Electr. Eng. **4** (2013) 39. https://doi.org/10.1016/j.compeleceng.2013.01.002
4 J. Wang, Z. Hual, and Huang: Engineering **4** (2017) 439. https://doi.org/10.1016/j.eng.2017.04.003
5 J. Yang, G. Zhang, Y. Mao, H. Li, and B. Yang: Adv. Mater. Res. **1065** (2015) 3696. https://doi.org/10.4028/www.scientific.net/amr.1065-1069.2239
6 L. Zhang, Z. Li, A, Li, and F. Liu: ISPRS J. Photogramm. Remote Sens. **138** (2018) 86. https://doi.org/10.1016/j.isprsjprs.2018.02.008
7 S. Malihi, M. J. Valadan Zoej, M. Hahn, M. Mokhtarzade, and H. Arefi: Remote Sens. Spatial Info. Sci. **1** (2016) 71. https://doi.org/10.5194/isprs-archives-xli-b3-71-2016
8 A. Dai, M. Niebner, M. Zollhofer, S. Izadi, and C. Theobalt: ACM Trans. Graphics **36** (2017) 76. https://doi.org/10.1145/3072959.3054739
9 Z. Zhang: Proc. IEEE Int. Conf. Computer Vision **1** (1999) 666. https://doi.org/10.1109/iccv.1999.791289
10 M. Camplani, T. Mantecon, and L. Salgado: IEEE Trans. Cybern. **43** (2013) 1560. https://doi.org/10.1109/tcyb.2013.2271112

About the Authors

Ningbo Jing received his B.S. and M.S. degrees in electrical engineering from Xi'an University of Science and Technology, China, in 2002 and 2006, respectively. From 2006 to 2009, he was an assistant engineer at Xi'an University of Science and Technology, China. Since 2010, he has been an engineer at Xi'an University of Science and Technology. His research interests are in MEMS, robots, and sensors. (153660621@qq.com)

Xianmin Ma received his B.S. and M.S. degrees in electrical engineering from Xi'an University of Mining and Technology, China, in 1982 and 1988, respectively, and his Ph.D. degree from China University of Mining and Technology University, in 1997. Since 1998, he has been a professor at Xi'an University of Science and Technology. His research interests are in robots and electric vehicle sensors. (1606826400@qq.com)

Wei Guo received his B.S. and M.S. degrees in mechanical design and production engineering from Xi'an University of Mining and Technology, China, in 1982 and 1988, respectively. From 1982 to 1998, he was an assistant professor at Xi'an University of Science and Technology, China. Since 1998, he has been a professor at Xi'an University of Science and Technology. His research interests are in automation control theory and coal mine machinery. (710699041@qq.com)

Mei Wang received her B.S. degree in automation from Changsha Railway College, China, in 1985 and her M.S. and Ph.D. degrees from Xi'an University of Science and Technology, China, in 1990 and 2006, respectively. Since 2006, she has been a professor at Xi'an University of Science and Technology. Her research interests are in artificial intelligence and sensors. (1309941881@qq.com)