

FENG JIN<sup>1</sup>, KAI ZHAN<sup>2</sup>, SHENGJIE CHEN<sup>3</sup>, SHUWEI HUANG<sup>4</sup>, YUANSHENG ZHANG<sup>5</sup>

## Ore extraction and analysis from RGB image and 3D Point Cloud

### Introduction

Ore is the most important product of underground mining. For mineral cracking system missions, cracking ore into small pieces or sweeping small pieces of ore down a winze is the objective. Nowadays, hand-handling cracking still plays a vital role in production procedure. With the development of a remote handling technique, a group-controlling crack device handled by professional workers is applied by an increasing number of mineral companies. This can improve production efficiency and reduce danger caused by workers who may handle cracking machines in the wrong way or break the production rule. With remote handling technique, the skillful worker can manipulate cumbersome machines from an office. This enables them to avoid noise and the risk of being hit by small pieces of ore. Therefore,

✉ Corresponding Author: Shuwei Huang; e-mail: shuwei.huang@hotmail.com

<sup>1</sup> BGRIMM Technology Group University of Science and Technology Beijing, China;

ORCID iD: 0000-0003-3978-4984; e-mail: jinfeng@bgrimm.com

<sup>2</sup> BGRIMM Technology Group, China; e-mail: zhankai@163.com

<sup>3</sup> BGRIMM Technology Group, China; e-mail: chenshengjie@163.com

<sup>4</sup> BGRIMM Technology Group, China; e-mail: shuwei.huang@hotmail.com

<sup>5</sup> BGRIMM Technology Group, China; e-mail: zhangyuansheng@163.com



© 2022. The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-ShareAlike International License (CC BY-SA 4.0, <http://creativecommons.org/licenses/by-sa/4.0/>), which permits use, distribution, and reproduction in any medium, provided that the Article is properly cited.

remote handling techniques will be widely applied in mineral production in the next few years. An automated crack system without people's participation is also becoming a reality. This paper discusses related work of the automated crack system. For an automated cracking system, different kinds of sensors are essential. For example, 3D industrial cameras, which can produce stereo information of environment, usually as a disparity map, a 3D point cloud. According to the principle of the device, 3D cameras can mainly be classified into three types – binocular cameras, ToF cameras (Gokturk et al. 2004; Salmerón et al. 2010; Foix et al. 2011), structured light cameras. Binocular cameras, based on two of the same cameras, generate a disparity map to get deep information. ToF camera, contains laser diodes and a ToF image sensor, and an output point cloud for further processing. Structured light cameras projects coded light to the surface of the object, and the use of the binocular camera's theory to calculate depth information, or make use of calibrated projected patterns and only a single camera is needed to compute depth from triangulation. A comparison of the three different 3D cameras is presented in Table 1.

Table 1. Comparison of three different types of 3D cameras

Tabela 1. Porównanie trzech różnych typów kamer 3D

| Property                    | Stereo vision                   | Structured light                | ToF cameras                |
|-----------------------------|---------------------------------|---------------------------------|----------------------------|
| Working Distance            | Limited by baseline             | Limited by baseline             | Scalable with light source |
| Depth Accuracy              | Low                             | High                            | Medium                     |
| Depth Map Resolution        | Limited by texture of scene     | Limited by light pattern        | Full resolution            |
| In-field Calibration Needed | Sometimes                       | Sometimes                       | No                         |
| Size                        | Increases with working distance | Increases with working distance | Compact                    |
| Cost                        | Low                             | High                            | Medium                     |

With regard to the texture of ore images, binocular cameras are not suitable for feature extracting and matching. For structured light cameras, due to the limitation of working distance, the precision decreases when the working distance is longer. The structured light camera performs the best between 1.0 meter and 4.0 meters. However, under realistic production circumstances, the working distance is mostly above 4.0 meters. Therefore, the structured light camera is not appropriate for an automated cracking system. The new-coming ToF cameras have the advantage of depth accuracy and good working distance. In addition, it can produce a high-quality 3D point cloud without feature extraction and key-point matching in subsequent procedures. Therefore, this paper chooses the ToF 3D camera for

this research. Additionally, a colorful industrial camera that provides RGB information is essential, but this can suffer from the harsh environment.

In general, the identification and segmentation of ore above the winze is very challenging, as ore has no regular shape, color, texture, or other features that can easily distinguish it from its environment. In addition, ore can be obscured by other ores or ore piles. Therefore, in this research, we consider the extraction algorithm based on the combination of RGB images and 3D point clouds. RGB image can provide texture and edge information, which is beneficial to segment ore regions from a 2D perspective. 3D point clouds can be used to calculate geometric information correspond with the segmented regions of RGB images according to the cross-calibration results of the RGBa and ToF cameras.

Many methods have been proposed to solve the problem of ore extraction. There have been a number of representative methods including watershed and its improved methods (Zhang and Jiang 2011; Zhang et al. 2011; Jin and Zhang 2018; Dong and Jiang 2013), threshold segmentation method (Zhang et al. 2019; Zhan and Zhang 2019). However, these traditional methods depend on time-consuming parameter-tuning processes to get ideal results and they are difficult to suitably fit under different conditions.

With the rapid development of deep learning theories, deep learning-based methods widely improve the performance of computer vision tasks, like image classification, object detection and semantic segmentation. For example, Faster RCNN (Ren et al. 2016) suits the object detection task. Liu et al. (Liu et al. 2020) applied U-Net to segment ore images. Meanwhile, Res-UNet has been used to optimize the segmentation masks. Xiao et al. (Xiao et al. 2020) applied a deformable convolution network (Dai et al. 2017) to replace the ordinary convolution operation and increase accuracy.

Our major contributions in this paper are as follows:

- ◆ we combine edge detection with an ordinary mean-shift algorithm to alleviate over-segment problem;
- ◆ we propose a novel method that calibrates the ToF camera with the RGB camera;
- ◆ we propose a new edge-based ore and ore pile classification method.

The rest of this paper is organised as follows:

- ◆ Section 1 introduces the method that segment ore with combination of embedded confidence edge detection algorithm and mean-shift algorithm;
- ◆ Section 2 mainly describes the cross-calibration procedure of the two cameras;
- ◆ Section 3 presents experimental results that algorithm classify ore and ore piles;
- ◆ finally, our conclusions are discussed in Section 4.

## 1. Automated ore extraction based on image segmentation algorithm

The ore image is initially segmented by a mean-shift algorithm (Comaniciu and Meer 2002; Cheng 1995; van den Boomgaard and van de Weijer 2002). Mean-shift is a method of non-parametric estimation, it operates a mode-seeking process through an iterative

procedure, which shifts each sample point to the average of the neighborhood sample within a D-dimensional feature space with a kernel function until it converges, and the objective is to find the maxima. Mean-shift has been widely used in various areas. For instance, pixel clustering, image segmentation (Pal and Pal 1993; Cheng et al. 2001), visual tracking, and discontinuity preserving smoothing.

The gradient-based feature space analysis method uses the gradients of the probability density function to find the maximum value. This method is complicated because the density probability needs to be estimated. Gradient-based methods first calculate the gradient and then the kernel is shifted by a specific length vector in the direction of the maximum increase of density. The magnitude is the step size that has to be appropriately chosen. The challenging task is how to choose a suitable step size because a small step size will slow down the convergence.

The mean-shift algorithm is an iterative method, which computes the mean-shift vector for the current set point, and then moves the set point to its mean-shift value. The stop criterion of computing the mean-shift vector process is that the algorithm reaches a specified convergence condition.

The mean-shift procedure consists of two steps:

- ◆ construction of probability density in some feature space;
- ◆ the mapping of each point to the maximum (mode) of the density which is closest to it.

We can describe mean-shift algorithm procedure as follows:

- ◆ Construct  $N$  5-dimensional points.
- ◆ Convert RGB image from RGB color space to  $LUV$  color space.
- ◆ Initialize  $y_{i,1} = c_i = (x_i^s, c_i^r)$ , where  $c_n^r$  is result of  $LUV$  representation.
- ◆ Judge whether the converge condition is fulfilled or not, if not, then calculate by Equation 1.  $y_{i,j+1}$  is the new position of the kernel window. If the converge condition is fulfilled, then set  $y_{i,conv} = y_{i,j+1}, z_i = (x_i^s, y_{i,conv}^r)$ .

$$y_{i,j+1} = \frac{\sum_{i=1}^n c_i g \left( \left\| \frac{y_{i,j} - c_i}{h} \right\|^2 \right)}{\sum_{i=1}^n g \left( \left\| \frac{y_{i,j} - c_i}{h} \right\|^2 \right)} \quad (1)$$

Calculate clusters  $\{C_p\}_{p=1,\dots,P}$  by merging those convergence points that are closer than  $h_r, h_s$ .

Set the label of mode sets  $L_i = \{p \mid z_i \in C_p\}$ .

Convert back the RGB value from the  $LUV$  result.

The weakness of the standard mean-shift algorithm is that the bandwidth parameter value deeply affects the segment results. If the selection of its value is not proper, an over-segment result will appear.



Fig. 1. The over-segmentation problem happens in the ore-segmenting procedure

Rys. 1. Problem nadmiernej segmentacji występujący w procedurze segmentacji rudy

In order to solve the over-segment problem, other methods should be proposed to work with the mean-shift algorithm together. In this research, the embedded confidence edge detection (Meer and Georgescu 2001) algorithm plays an important role in improving the segmentation result. The output of mean-shift is the number of regions. Each region has the same label, and the RGB value in each segmented area is the same. The boundary extraction of each segmented region is essential for the subsequent steps, the objective of which is to merge over-segment regions that belong to the whole ore. The results of the embedded confidence edge detection method are distributed near the boundary of the segmentation area.

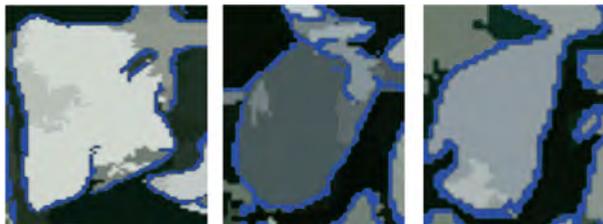


Fig. 2. Use of surrounding edges of region boundary to merge adjacent regions

Rys. 2. Użycie otaczających krawędzi granicy regionu w celu scalenia sąsiedniego regionu

The distribution of edge points around the region boundary can determine whether adjacent regions can be merged. Traditional edge detection based on the gradient magnitude changes. In order to calculate the edge magnitude, a gradient operator should be introduced to obtain the edge amplitude and direction. Applying gradient magnitude with the non-maximal and suppression method to get final result. In order to detect weak edges and restraint

noise, an embedded confidence algorithm based on gradient direction combined with gradient size is proposed to calculate the confidence.

The procedure of the embedded confidence edge detection algorithm is as follows:

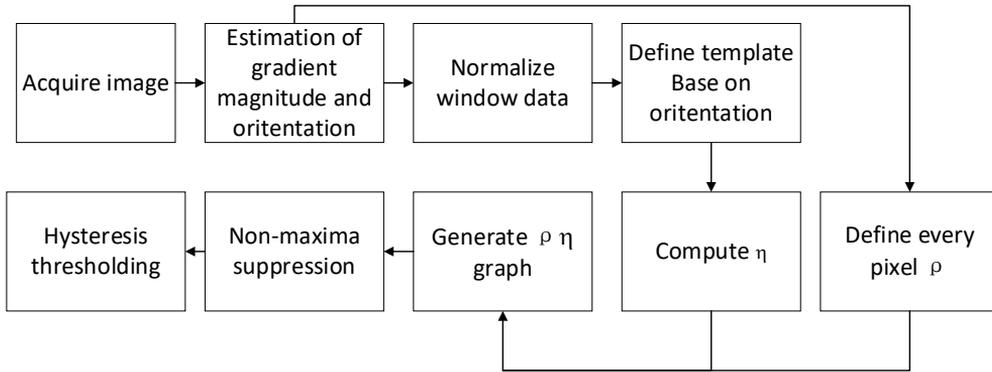


Fig. 3. The flowing chart of embedded edge detection algorithm

Rys. 3. Schemat blokowy wbudowanego algorytmu wykrywania krawędzi

According to edge distribution around the boundaries of adjacent regions, this research set conditions to judge whether to merge regions or not.

Setting  $n_w$  represents the sum of the boundary, and defining  $n_e$  as the surrounding edges of the boundary between adjacent regions, then calculating  $p_r$  who represents the proportion of edge points comparative to the whole boundary numbers, the  $p_r$  is calculated as:

$$p_r = n_e/n_w \quad (2)$$

Given threshold  $t_c$ , if  $p_r > t_c$ , then merge adjacent regions and set the same label to all merged regions. If  $p_r \leq t_c$ , then do not merge region, and the algorithm keeps iterating.

## 2. Cross-calibration of the ToF and RGB cameras

In Section 2, we discussed ore RGB image segment processing. In order to analyze the ore image from the Geometric perspective, we should combine RGB image analysis with 3D information of the ore. As previously mentioned, the ToF camera is applied to acquire the point cloud. In this section, we mainly discuss how to transform the ToF camera coordinate to the RGB camera coordinate. For some specific RGB cameras, manufactures integrate this function within the camera. In this research, we manually unify the coordinates of the two cameras, and present an accurate and practical method for joint calibration of ToF cameras and color cameras. Firstly, the model of the RGB camera used in this research is

TRI032S-CC manufactured by LUCID Vision Labs, the 3D ToF camera used in this research is Helios manufactured by LUCID Vision Labs. The specifications of these two cameras are presented in Table 2.

Table 2. Specification of two cameras

Tabela 2. Specyfikacja dwóch kamer

| Property          | TRI032S-CC          | Helios2             |
|-------------------|---------------------|---------------------|
| Sensor            | Sony IMX265 CMOS    | Sony IMX556PLR      |
| Resolution        | 2048 x 1536, 3.2 MP | 640 x 480, 0.3MP    |
| Frame Rate        | 35.4fps@3.2MP       | 30fps               |
| Digital Interface | Power over Ethernet | Power over Ethernet |
| Lens              | C-Mount             | Interior            |

We should place two cameras as close together as possible. In order to reduce the amount of occlusion between the cameras, a common placement method is mounting the RGB camera on the side or on the top of the ToF camera. In this research, we use top-down placement as shown in Figure 7, and the front view of the camera combination is shown in Figure 8.

The calibration procedure requires a standard calibration plate, which is consisted of many black and white pattern circles and contains 10–20 images. For some low-resolution ToF cameras, the accuracy does not perform well in calibration. Since some ToF cameras only provide



Fig. 4. Top-down placement

Rys. 4. Sposób umieszczenia od góry

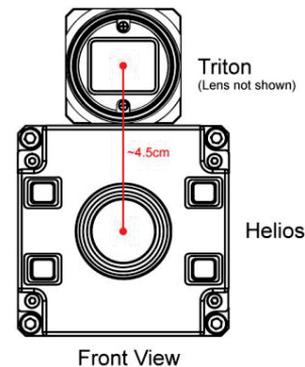


Fig. 5. Front view of placement

Rys. 5. Widok z przodu miejsca ustawienia

depth images, the corner in depth image can't be detected very well. As a consequence, we use the Lucid ToF camera to acquire intensity image from which we detect circle corner and then calculate depth information. Next, we introduce the perspective projection model.

### 2.1. The perspective projection model

In order to estimate the 3D pose transformation between the ToF camera and the RGB camera, we need numerous 3D-to-2D correspondences between  $n$  specified points  $M_1, \dots, M_n$  where  $M_i = [X, Y, Z]^T$  are formed in a Euclidean world coordinate system and their corresponding 2D projections points  $m_i, \dots, m_n$ , and  $m_i = [u, v]^T$  in the image plane. Thus, the defined projection can be expressed within the equation:

$$sp_i = MP_i \quad (3)$$

Where  $s$  is a scale factor,  $p_i = [u, v, 1]^T$  and  $P_i = [X, Y, Z, 1]^T$  are the homogeneous coordinates of points  $p_i$  and  $P_i$ , and  $M$  is a  $3 \times 4$  projection matrix.

It is known that  $M$  is defined by a scale factor which also depends on 11 parameters. The perspective matrix can be expressed as:

$$M = K[R/T] \quad (4)$$

Where  $K$  is  $3 \times 3$  matrix which contains inner camera property, such as focal length, scale factor, and optical center point. And,  $K$  can be expressed as the following patterns:

$$K = \begin{bmatrix} f/d_x & 0 & u_0 \\ 0 & f/d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

Where  $1/d_x$  and  $1/d_y$  are the number of pixels per unit in the horizontal and vertical directions of the inner camera imaging chip.  $[u_0, v_0]^T$  represents the principal point coordinates, which is the intersection of the optical axis and the image plane.

We defined  $[R/T]$  as a  $3 \times 4$  matrix that corresponds to the Euclidean transformation from a world coordinate system to the camera coordinate system. In fact, this matrix is the horizontal concatenation of the rotation matrix and the translation vector which is often referred to as the camera pose.

$$[R | t] = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_1 \\ R_{21} & R_{22} & R_{23} & t_2 \\ R_{31} & R_{32} & R_{33} & t_3 \end{bmatrix} \quad (6)$$

In Figure 4, we depict the camera coordinate system, the world coordinate system, and the image plane.  $(X_W, Y_W, Z_W)$  is the world coordinate system,  $(X_C, Y_C, Z_C)$  is the camera coordinate system,  $(u, v)$  is the imaging plane coordinate system,  $M(X, Y, Z)$  is a point in the camera coordinate system, and  $m_r(x_r, y_r)$  is the realistic physical coordinate of  $M(X, Y, Z)$  in the imaging plane,  $m_i(x_i, y_i)$  is the ideal physical coordinate of  $M(X, Y, Z)$  in the imaging plane.

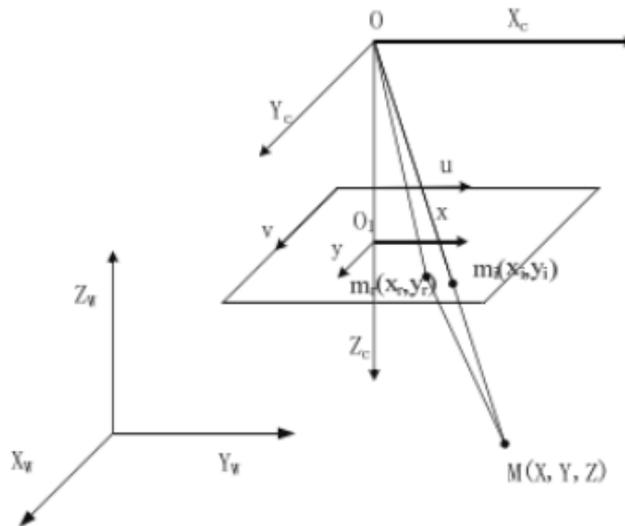


Fig. 6. The camera model

Rys. 6. Model aparatu

## 2.2. The procedure of the cross-calibration the ToF and RGB cameras

Lots of research and methods exist in relation to the color camera calibration task. The fundamental method is proposed in the literature (Zhang 2000), several images containing a checker pattern are acquired through the RGB camera, corners of the checker on the board are detected in the calibration procedure and are then used to calculate the intrinsic and extrinsic parameters of the RGB camera through a non-linear optimization method. Many iterative solutions to the over-constrained problem with  $n > 3$  point correspondences also exist.

The main objective of pose estimation between the two cameras is to find the perspective projection matrix  $P$  which projects the 3D points  $M_i$  on  $m_i$ , given a set of  $n$  correspondences. We can rewrite that in terms of  $PM_i = m_i$  for all  $i$ .

The camera pose estimation from 3D to 2D points is a fundamental problem in geometric computer vision area that has already been solved. The objective is to estimate the six degrees of freedom of the camera pose and the camera calibration parameters: the focal length, the principal point, the aspect ratio and the skew. An approach to solve the problem

using a minimum of six pairs of correspondences can be achieved by using the direct linear transform (DLT) algorithm (Hartley and Zisserman 2003). While the DLT algorithm requires the camera parameters, numerous simplifications to the problem have been proposed to improve the solution accuracy and form a large set of new different algorithms. A variant of the previous algorithm is called the Perspective- $n$ -Point problem, which assumes that the intrinsic parameters of the camera are known. In its minimal version, only three point correspondences are needed to recover the camera pose (Gao et al. 2003). There also exist many iterative solutions to the over-constrained problem with  $n > 3$  point correspondences.

When the points number involved in solving PnP only equals 3, the PnP problem is treated as a minimal form of PnP and can be solved with just three point correspondences. However, with just three point pairings, P3P generates up to four real, geometrically feasible solutions. For low noise levels, a fourth correspondence can be used to erase ambiguity. The details for the problem are as follows, and the P3P problem in the geometrical pattern is shown in Figure 7.

Let  $P$  be the center of projection for the camera,  $A, B, C$  be 3D world points with corresponding images points  $u, v$ , let  $a = |PA|$ ,  $b = |PB|$ ,  $c = |PC|$ ,  $\theta_{ab} \angle APB$ ,  $\theta_{ac} \angle APC$ ,  $\theta_{bc} \angle BPC$ ,  $p = 2\cos\theta_{bc}$ ,  $q = 2\cos\theta_{ac}$ ,  $r = 2\cos\theta_{ab}$ ,  $R_{ab} = |AB|$ ,  $R_{ac} = |AC|$ ,  $R_{bc} = |BC|$ , this generates triangles  $PAC$ ,  $PAB$ ,  $PBC$ .

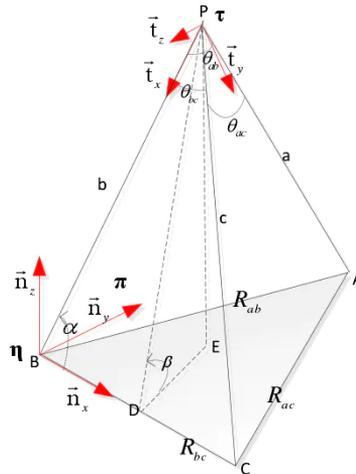


Fig. 7. The P3P problem

Rys. 7. Problem P3P

We obtain the following equation system for P3P:

$$\begin{cases} a^2 + b^2 - abr - R_{ab}^2 = 0 \\ a^2 + c^2 - acq - R_{ac}^2 = 0 \\ b^2 + c^2 - bcp - R_{bc}^2 = 0 \end{cases} \quad (7)$$

We apply the zero decomposition method (Wu 1986), which is recognized as a fundamental method to solve systems of algebraic equations to handle the P3P equation system. By solving the P3P system, four geometrically feasible real solutions of R and T can be obtained.

After acquiring R and T, in the last step we will project all 3D points from the ToF to the RGB camera coordinate system. Some specific ToF cameras can generate intensity images which do not contain color information. Values of the intensity image represent reflective intensities within some range. For the same calibration boards, the intensity image can also be used to detect circles in it. Every pixel in the intensity image has a depth value if the reflection of the object material is strong enough. Therefore, the RGB camera and the ToF camera can capture a chessboard at the same position. Then, steps of cross-calibration are listed as follows:

- ◆ Prepare the calibration board and get 10–20 images from the RGB camera.
- ◆ Detect circles on the board in the RGB image and save all center coordinates of circles.
- ◆ Calculate extrinsic and intrinsic parameters and the distortion coefficient of the RGB camera.
- ◆ Acquire an intensity image that contains the calibration board and get a corresponding 3D point cloud of this scene.
- ◆ Snap an RGB image at the same position as the ToF camera.
- ◆ Apply the PnP method to calculate the transformation from the ToF 3D camera to the RGB camera, the output of this procedure is the rotation vector and the translation vector.

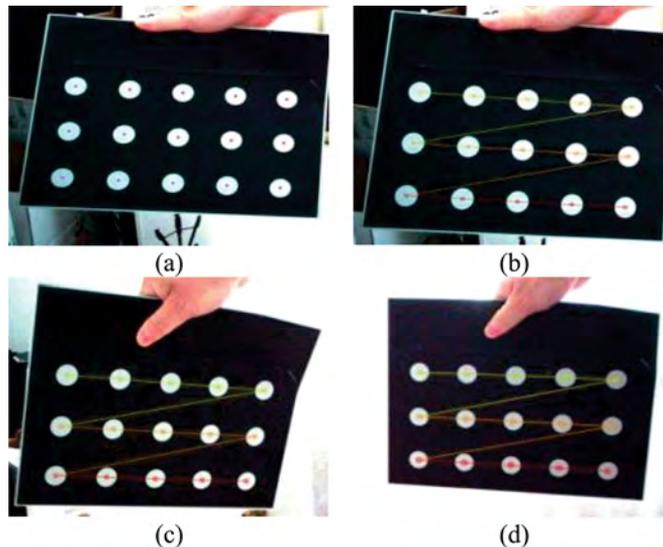


Fig. 8. (a) circles detected result using the Blob analysis method from different view to estimate the camera matrix. (b), (c) and (d) detected circles connected from left to right then from top to bottom

Rys. 8. (a) okręgi wykryły wynikiem metodą analizy Blob z innego punktu widzenia w celu oszacowania macierzy aparatu. (b) (c) (d) wykryte okręgi połączone od lewej do prawej, a następnie od góry do dołu

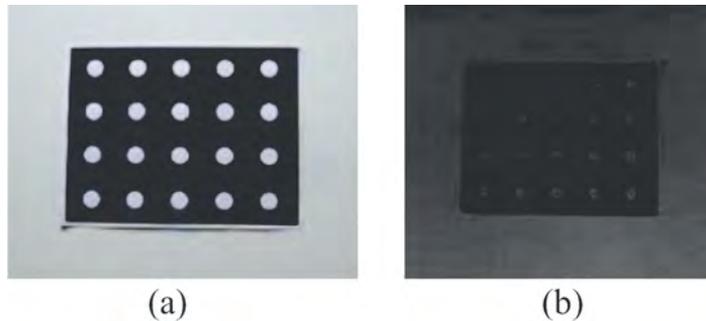


Fig. 9. (a) Chessboard image captured by the RGB camera after calibration.  
(b) Chessboard intensity image captured by the ToF camera at the same position as (a)

Rys. 9. (a) Obraz szachownicy przechwycony przez kamerę RGB po kalibracji.  
(b) Obraz intensywności szachownicy przechwycony przez kamerę ToF w tej samej pozycji co (a)

In Figure 9, the chessboard is at the same position under two different camera views. The RGB image and intensity image are captured by the Triton and Helios cameras respectively, then, we detect circles on the chessboard to acquire all circles in the 2D position in the RGB camera coordinate system and the 3D position in the TOF camera coordinate system.

Finally, all 2D-3D coordinate pairs are input into the Solve PnP algorithm to calculate the rotation and transition transformation.

### 3. Experimental results

Since the segmentation results of RGB image can only represent the two-dimensional area of the scene, we cannot judge the classification of different regions. For instance, features of ore image above the winze are similar to the ground; in order to distinguish all ore and piles of the ore regions from the scene, we should make use of a corresponding 3D point cloud according to the cross-calibration result. Firstly, we put some labels on the plane of the mine, and calculate the XYZ value of these labels, then fit the plane of the ore mine winze.

As shown in Figure 10, we use the 3D coordinates of nine red rectangle labels to fit the plane of the winze. We then calculate the height of each segmented area above the winze plane to determine whether it is an ore area. When the height of the area exceeds a given threshold, it can be judged as ore or ore piles. A number of RGB images and corresponding 3D point clouds were obtained from the winze of an underground mine. There exists ore and piles of small pieces of ore. In order to work with an automated cracking machine, the classification of two kinds of ore is essential. Ore pile is not suitable for a cracking device to crack. Owing to this, no accurate cracking point can be provided to crack them down.

In this research, we propose an edge-based classification method. Given different edge length parameters, we calculate the edge distribution around the merged region. If there are

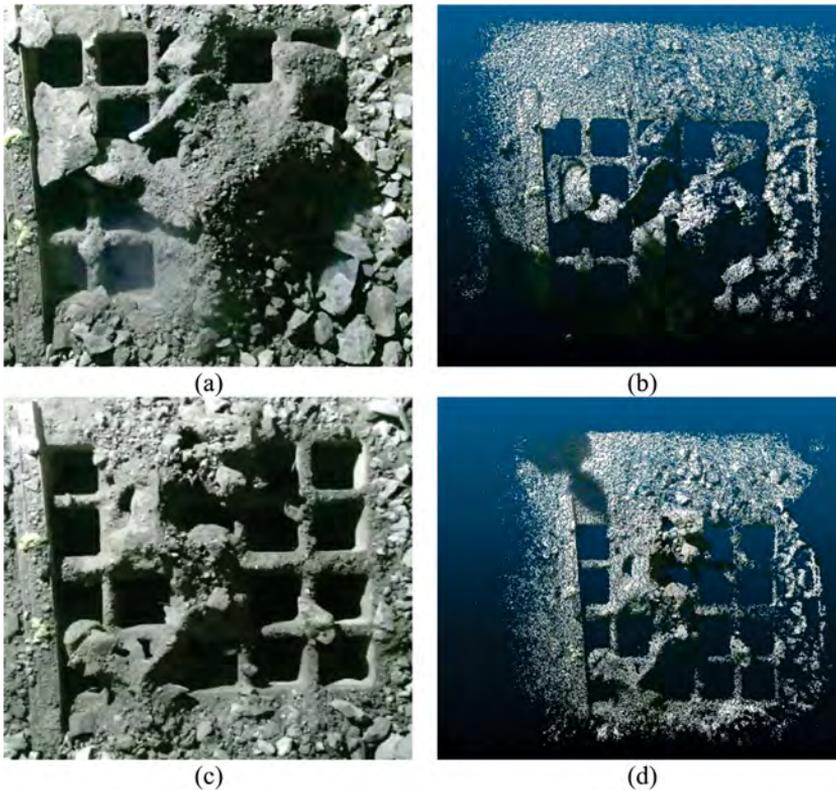


Fig. 10. Use of the coordinate transformation of two camera to overlay the RGB value on the point cloud (a) original image; (b) corresponding point cloud of (a); (c) original image; (d) corresponding point cloud of (c)

Rys. 10. Zastosowanie transformacji współrzędnych dwóch kamer do nałożenia wartości RGB na chmurę punktów; (a) oryginalny obraz; (b) odpowiednia chmura punktów z (a); (c) oryginalny obraz i (d) odpowiednia chmura punktów (c)

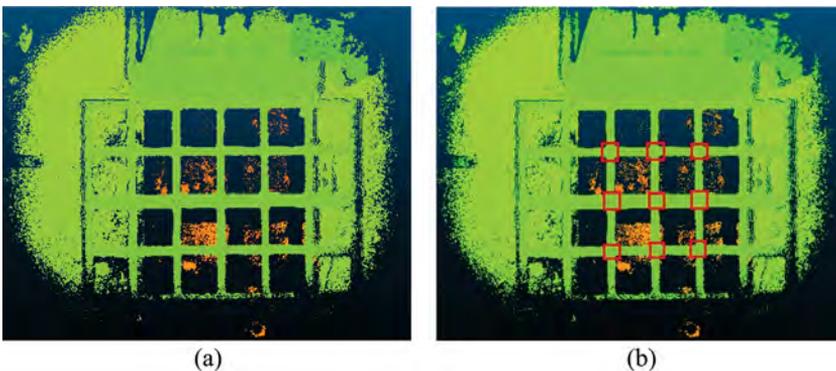


Fig. 11. The empty ore winze plane point cloud. The left image (a) is empty winze. Labels on the winze plane appeared in right image (b)

Rys. 11. Pusta chmura punktów płaszczyzny rudy winze. Lewy obraz (a) jest pustym winze. Etykiety na płaszczyźnie winze pojawiły się na obrazku (b)

short sides with a length of less than 50, the area is recognized as be a mine pile. If not, the region is recognized as ore. Finally, according to the cross-calibration results of the RGB camera and the ToF camera, the plane area of the RGB image is given, and a 3D point cloud of this region can be acquired, which is crucial for the subsequent cooperating with the cracking machine.

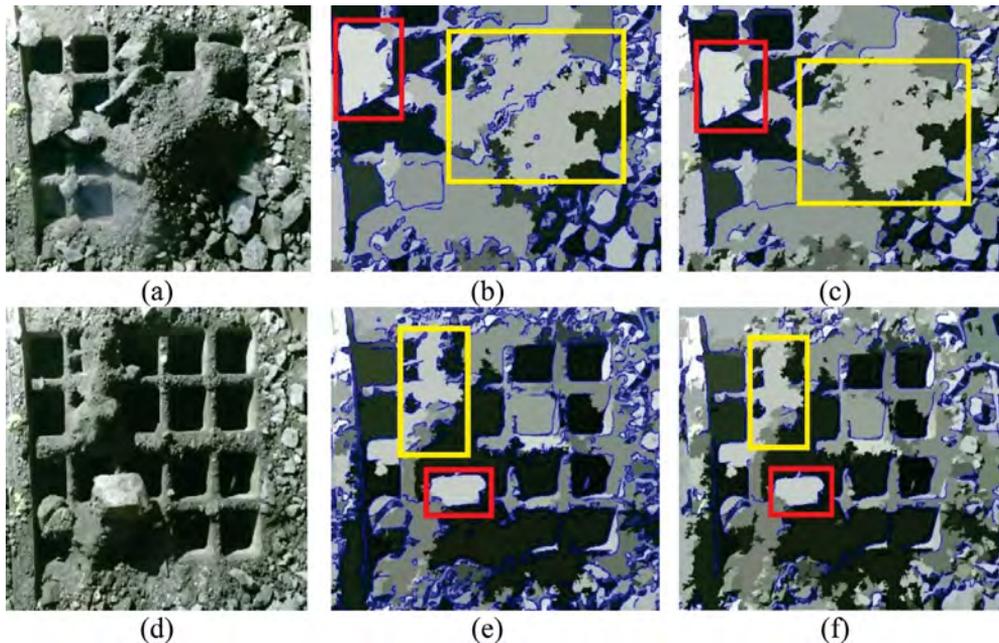


Fig. 12. Applying the edge length parameter to classify types of ore (a) original image; (b) corresponding edge detection result, length parameter equals 10; (c) corresponding edge detection result, length parameter equals 50; (d) original image; (e) corresponding edge detection result, length parameter equals 10; (f) corresponding edge detection result, length parameter equals 50

Rys. 12. Zastosowanie parametru długości krawędzi do klasyfikacji rodzajów rudy; (a) oryginalny obraz; (b) odpowiedni wynik wykrywania krawędzi, parametr długości równy 10; (c) odpowiedni wynik wykrywania krawędzi, parametr długości równy 50; (d) oryginalny obraz; (e) odpowiedni wynik wykrywania krawędzi, parametr długości równy 10; (f) odpowiedni wynik wykrywania krawędzi, parametr długości równy 50

From Figure 11, we can obviously see that edge detection around ore and the ore pile region boundaries do not disappear when the edge detection parameter equals 10. However, when the edge length parameter is set to 50, the edges nearly vanish. By contrast, the edge points of the regions in the red rectangles do not nearly vanish, and it can be classified into the ore type, and the regions in the yellow rectangles can be classified into the ore pile.

## Conclusions

In this paper, a new ore image extraction method is proposed that combines the mean shift method and the embedded confidence edge detection algorithm. According to the edge distribution around region boundaries to judge whether to merge adjacent regions. If the ratio of the edge point to the region boundary point exceeds a certain level, the two adjacent areas do not need to be merged. Instead, we should merge these two regions.

Cross-calibration reveals the relationship between the RGB camera and the ToF camera, which are based on a classic method using a calibration board. The first step is to get the internal and external parameters of the RGB camera and then the ToF camera to calculate depth information based on intensity image. Finally, the PnP method is used to project the three-dimensional information of all the circles on the aircraft from the ToF to the RGB camera coordinate system to obtain the transformation of the two cameras, usually in the form of a rotation vector and a translation vector.

From the realistic production environment, there exists ore mixed with ore piles. In order to make the cracking machine work efficiently, it is important to accurately classify these two ore types. In this research, the edge length parameter is used to determine the attribution of this type of area. Due to the sharpness of the ore, the mining area has fewer short sides. However, the ore pile region has more short edges because of the irregularity of the region boundary.

In conclusion, due to the similarity of texture and color, it's difficult to segment the ore region from the 2D image perspective. From the 3D view, the boundary of the ore can't be acquired easily by only a point cloud. Therefore, we consider combining a 2D RGB image together with a point cloud to extract the ore region and integrating the embedded edge detection with mean-shift to solve the over-segment problem in the basic segmentation process. In the future, with the rapid development of sensors, we can predict that only one 3D-related device will be needed to solve the problem of ore segmentation.

*This work was jointly supported by the Major Science and Technology Innovation Project of Shandong Province (No. 2019SDZY05) and the Scientific Research Fund of BGRIMM Technology Group (No. 02-2035).*

## REFERENCES

- Cheng et al. 2001 – Cheng, H.D., Jiang, X.H., Sun, Y. and Wang, J. 2001. Color image segmentation: Advances and prospects. *Pattern Recognition* 34(12), pp. 2259–2281, DOI: 10.1016/S0031-3203(00)00149-7.
- Cheng, Y. 1995. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(8), pp. 790–799, DOI: 10.1109/34.400568.
- Comaniciu, D. and Meer, P. 2002. Mean shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Analysis and Machine Intelligence* 24(5), pp. 603–619, DOI: 10.1109/34.1000236.
- Dai et al. 2017 – Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H. and Wei, Y. 2017. Deformable convolutional networks. *Proceedings of the IEEE international conference on computer vision*. pp. 764–773, DOI: 10.1109/ICCV.2017.89.

- Dong, K. and Jiang, D.-L. 2013. Ore image segmentation algorithm based on improved watershed transform. *Computer Engineering and Design* 34, pp. 899–903
- Foix et al. 2011 – Foix, S., Alenya, G. and Torras, C. 2011. Lock-in time-of-flight (ToF) cameras: A survey. *IEEE Sensors Journal* 11(9), pp. 1917–1926, DOI: 10.1109/JSEN.2010.2101060.
- Gao et al. 2003 – Gao, X.S., Hou, X.R., Tang, J. and Cheng, H.F. 2003. Complete solution classification for the perspective-three-point problem. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 25(8), pp. 930–943, DOI: 10.1109/TPAMI.2003.1217599.
- Gokturk et al. 2004 – Gokturk, S. B., Yalcin, H. and Bamji, C. 2004. A time-of-flight depth sensor-system description, issues and solutions. 2004 conference on computer vision and pattern recognition workshop. IEEE, pp. 35–35, DOI: 10.1109/CVPR.2004.17.
- Hartley, R. and Zisserman, A. 2003. *Multiple view geometry in computer vision*. DOI: 10.1017/CBO9780511811685.
- Jin, X. and Zhang, G. 2018. Ore Impurities Detection Based on Marker-Watershed Segmentation Algorithm. *Computer Science and Application* 8, pp. 21–29, DOI: 10.12677/csa.2018.81004.
- Liu et al. 2020 – Liu, X., Zhang, Y., Jing, H., Wang, L. and Zhao, S. 2020. Ore image segmentation method using U-Net and Res\_Unet convolutional networks. *RSC Advances* 10, pp. 9396–9406, DOI: 10.1039/C9RA05877J.
- Meer, P. and Georgescu, B. 2001. Edge Detection with Embedded Confidence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, pp. 1351–1365, DOI: 10.1109/34.977560.
- Pal, N. R. and Pal, S. K. 1993. A review on image segmentation techniques. *Pattern Recognit* 26(9), pp. 1277–1294, DOI: 10.1016/0031-3203(93)90135-J.
- Ren et al. 2016 – Ren, S., He, K., Girshick, R. and Sun, J. 2016. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence* 39, pp. 1137–1149.
- Salmerón et al. 2010 – Salmerón, S. F., Ribas, G. A. and Genís, C. T. 2010. *Exploitation of time-of-flight (ToF) cameras*. Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Tech. Rep. IRI-TR-10-07.
- Van den Boomgaard, R. and van de Weijer, J. 2002. On the equivalence of local-mode finding, robust estimation and mean-shift analysis as used in early vision tasks. Object recognition supported by user interaction for service robots. *IEEE*, pp. 927–930, DOI: 10.1109/ICPR.2002.1048187.
- Wu, W. 1986. Basic Principles of Mechanical Theorem Proving in Elementary Geometries. *Journal of Automated Reasoning* 2, pp. 221–252, DOI: 10.1007/BF02328447.
- Xiao et al. 2020 – Xiao, D., Liu, X., Le, B.T., Ji, Z. and Sun, X. 2020. An ore image segmentation method based on RDU-Net model. *Sensors* 20(17), DOI: 10.3390/s20174979.
- Zhan, Y. and Zhang, G. 2019. An improved OTSU algorithm using histogram accumulation moment for ore segmentation. *Symmetry* 11(3), DOI: 10.3390/sym11030431.
- Zhang et al. 2011 – Zhang, G.Y., Liu, G.Z. and Zhu, H. 2011. Segmentation algorithm of complex ore images based on templates transformation and reconstruction. *International Journal of Minerals, Metallurgy, and Materials* 18(4), pp. 385–389, DOI: 10.1007/s12613-011-0451-8.
- Zhang et al. 2019 – Zhang, J.L., Sun, S.S. and Qin, S.Q. 2019. Ore image segmentation based on optimal threshold segmentation of genetic algorithm. *Science Technology and Engineering* 19, pp. 105–109.
- Zhang, W. and Jiang, D. 2011. The marker-based watershed segmentation algorithm of ore image. *IEEE 3<sup>rd</sup> International Conference on Communication Software and Networks*, pp. 472–474, DOI: 10.1109/ICC-SN.2011.6014611.
- Zhang, Z. 2000. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence* 22(11), pp. 1330–1334, DOI: 10.1109/34.888718.

**ORE EXTRACTION AND ANALYSIS FROM RGB IMAGE AND 3D POINT CLOUD****Key words**

ore image, 3D point cloud, embedded confidence edge detection, mean-shift, cross-calibration

**Abstract**

Based on the theory of computer vision, a new method for extracting ore from underground mines is proposed. This is based on a combination of RGB images collected by a color industrial camera and a point cloud generated by a 3D ToF camera. Firstly, the mean-shift algorithm combined with the embedded confidence edge detection algorithm is used to segment the RGB ore image into different regions. Secondly, the effective ore regions are classified into large pieces of ore and ore piles consisting of a number of small pieces of ore. The method applied in the classification process is to embed the confidence into the edge detection algorithm which calculates edge distribution around ore regions. Finally, the RGB camera and the 3D ToF camera are calibrated and the camera matrix transformation of the two cameras is obtained. Point cloud fragments are then extracted according to the cross-calibration result. The geometric properties of the ore point cloud are then analysed in the subsequent procedure.

**WYDOBYCIE I ANALIZA RUDY Z OBRAZU RGB I CHMURY PUNKTÓW 3D****Słowa kluczowe**

obraz rudy, chmura punktów 3D, wbudowane wykrywanie krawędzi ufności, zmiana średniej, kalibracja krzyżowa

**Streszczenie**

W oparciu o teorię widzenia komputerowego zaproponowano nową metodę wydobywania rudy z podziemnych kopalń. Jest to połączenie obrazów RGB zebranych przez kolorową kamerę przemysłową oraz chmury punktów wygenerowanej przez kamerę 3D ToF. Po pierwsze, algorytm zmiany średniej w połączeniu z wbudowanym algorytmem wykrywania krawędzi ufności służy do segmentacji obrazu rudy RGB na różne regiony. Po drugie, efektywne regiony rud są podzielone na złoża rudy o dużych rozmiarach i zwałowiska składające się z małej ilości rudy. Metodą stosowaną w procesie klasyfikacji jest określenie ufności w algorytmie wykrywania krawędzi, który oblicza rozkład krawędzi wokół regionów rudnych. Wreszcie, kamera RGB i kamera 3D ToF są skalibrowane i uzyskuje się transformację macierzy obu kamer. Następnie, fragmenty chmury punktów są wyodrębniane zgodnie z wynikiem kalibracji krzyżowej. W kolejnej procedurze przeanalizowano właściwości geometryczne chmury punktów rudy.

