Acta Technica 62, No. 3B/2017, 483–496

Kinect-based identification method for parts and disassembly track[1]

Zhang Zhijia[2,3], Wei Xin[2,3], Zhou Ziqiang[3], Wang Tingting[2], He Xin[2]

Abstract. A novel method is proposed for parts and disassembly track identification based on Kinect for robotic disassembly. Firstly, Kinect sensor is used to obtain color and depth image of the parts. According to the correspondence of color and depth image, parts can be identified, and the spatial location of parts can be located. Secondly, Kinect is used to get depth images of disassembly personnel and disassembly tools, and the binary depth image is refined to extract the trajectory features of disassembly tools. Finally, the trajectory identification is realized through training trajectory samples of the hidden Markov model. The actual scene identification test for robotic disassembly shows that the proposed algorithm can effectively identify typical parts and get the movement trajectory of disassembly tools.

Key words. Kinect sensor, parts identification, trajectory identification, robotic disassembly.
.

1. Introduction

The disassembly is the prerequisite to the recovery of typical parts, and the remanufacturing cannot be carried out without the disassembly. At present, some waste mechanical products in China are completed by some recovery companies, but the disassembly task still mainly focuses on the manual disassembly. With the scientific development, the autonomous robot plays an important role in the manufacturing industry, which is not yet used in the disassembly of waste mechanical products, and the main reason is that the autonomous robot cannot realize automatic identification and disassembly of typical parts on waste mechanical products.

The machine vision technology is widely used in military affairs, crop quality

[1]Key Construction Laboratory Fund of Mechanical Product Recycling Technology, Jiangsu Province,China (KF1508), National Natural Science Foundation of China (61540069).

[2]College of Software, Shenyang University of Technology, Shenyang, 110020, China

[3]Jiangsu Key Laboratory of Recycling and Reuse Technology for Mechanical and Electronic Products, Changshu, 215500 China

detection, face recognition, fingerprinting and medical image detection with its advantages of non-contact, wide spectral response range, super-long standby time, location, measurement and defect detection [1]–[3].

In the machine vision, most of image acquisition devices have characteristics such as complicated manufacturing process, high accuracy and expensive price. With Bumblebee2 as an example, this camera is of high price and large volume. However, the Microsoft Kinect camera is widely used in fields including 3D reconstruction, object tracking and posture recognition [4]–[6] because it is simple, cheap and convenient.

In this paper, with automatic robotic disassembly oriented relevant application, identification and location of typical parts are studied based on Kinect, which is then used for extraction and identification of the trajectory of the disassembly tools.

2. Kinect device and data acquisition

Kinect was XBOX360 kinect peripheral formally launched by Microsoft in 2010. Kinect has three cameras in total (Fig. 1), in which the middle lens is RGB color camera, and two lens on left and right side are infrared transmitter and infrared CMOS camera respectively. The drive motor is equipped at the bottom of Kinect, and array microphone system is built-in on both sides, which is used for the speech recognition. Kinect system [7] outputs color image and depth image of 640×480 by RGB color camera and depth sensor (Fig. 2), and any point in Kinect vision range can acquire its corresponding color information and depth information.

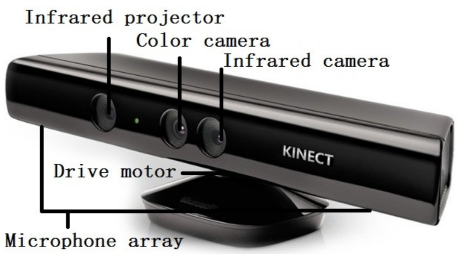

Fig. 1. Kinect sensor

Kinect has the skeleton tracking technology, which can determine each part of human body through establishing coordinates of each joint of human body (such as hand, head and body) by processing of depth data, and can determine the spatial location of them relative to Kinect. Based on this feature, in the disassembly process, Kinect can be used to track the human skeleton.

Kinect system expresses a skeleton through 20 joints. When a person enters into the Kinect vision range shown in Fig. 3, Kinect will find 20 skeleton joints of the person, which can be expressed in the form of coordinates.

Fig. 2. Kinect image: left–depth image, right–color image

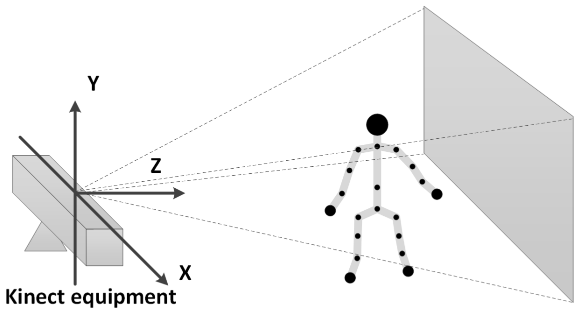

Fig. 3. Human skeleton in Kinect scene

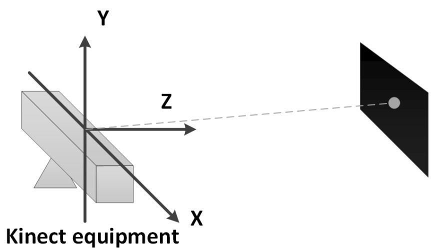

Fig. 4. Sketch map of black wood

3. Identification and location of typical parts

As shown in Fig. 2, the color image contains more image details than the depth image, which can satisfy the requirements for the identification of some parts with complex structure. However, the color image only extracts the coordinate information in the image, while the advantage of the depth image is able to extract 3D spatial information of parts, reaching the requirements for identification and location of parts. Based on the above feature, depth image and color image can be used at the same time for the identification, and 3D spatial position of parts can be determined.

3.1. Image transformation

Because of Kinect device itself, as shown in Fig. 2, the objects in Kinect depth image and color image are of different sizes, and the figure in the depth image is larger. After the image transformation, color image and depth image coincide, and the identification of the coincided color image is carried out, which facilitates the subsequent location of parts.

A regular square black wood plane is taken as the object, which is scanned with Kinect, to acquire depth image and color image. Fig. 4 gives a schematic diagram of black wood scanned with Kinect, and Fig. 5 gives the acquired depth image and color image.

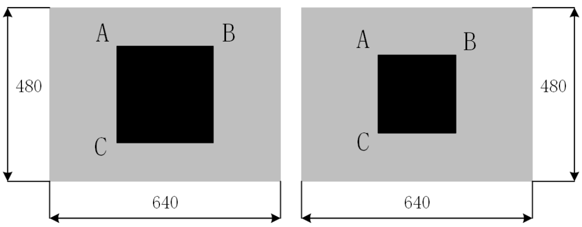

Fig. 5. Schematic diagram of Kinect image: left–depth image, right–color image

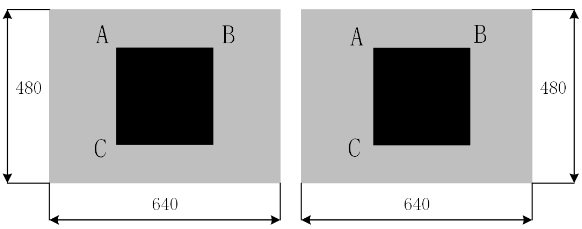

Fig. 6. Schematic diagram after affine transformation: left–depth image, right–affine transformed color image

As shown in Fig. 5, 2D coordinates $(x_{\mathrm{a}}, y_{\mathrm{a}})$, $(x_{\mathrm{b}}, y_{\mathrm{b}})$ and $(x_{\mathrm{c}}, y_{\mathrm{c}})$ of three angles - A, B and C of black part in the depth image are extracted. The coordinates $\left(x_{\mathrm{a}}^{'}, y_{\mathrm{a}}^{'}\right)$, $\left(x_{\mathrm{b}}^{'}, y_{\mathrm{b}}^{'}\right)$ and $\left(x_{\mathrm{c}}^{'}, y_{\mathrm{c}}^{'}\right)$ of three angles - A, B and C of black part in the color image are extracted, and coordinates of three pairs of points are substituted into the matrix

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x^{'} \\ y^{'} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \tag{1}$$

Simultaneous six equations are used to solve six unknown number to obtain the affine transformation matrix of 2×3. The affine transformation of the color image is carried out with the affine transformation matrix, and Fig. 6 gives the schematic diagram of contrast after the transformation. As shown in Fig. 6, after the transformation, the positions of A, B and C in the color image are coincided with the positions of A, B and C in the depth image.

3.2. Recognition of parts

After the affine transformation, color image and depth image coincide, and the identification of the coincided color image is carried out, to facilitate the subsequent location of parts. With the hex nut as an example, the color image of parts is extracted with Kinect. The affine transformation is carried out for the color image according to affine transformation matrix, and Fig. 7 gives the contrast diagram before and after the affine transformation.

Fig. 7. Affine transformation: left–color image, right–affine transformed image

As shown in Fig. 7, after the affine transformation, scales of color image and depth image are consistent. After the completion of affine transformation of the image, the color image is transformed to gray image.

According to the gray image, the correlation coefficient matching method is used for the identification. This method is used to find out a measurement, to judge whether the template matches best with the image searched in the search graph. The correlation coefficient method is a relatively good method to express the matching measurement. After matching, parts in the image are identified according to the similarity.

3.3. Calculation of part pose

The poses of typical parts are coordinates in the spatial coordinate system, namely, world coordinate, while the previous section is only to obtain the coordinate in the image with matching algorithm, which is difficult to locate. If spatial poses of typical parts are required, the correspondence of color image and depth image can be used. Fig. 8 gives the schematic diagram of part location.

As shown in Fig. 8, the coordinate of the central point P of typical parts is extracted through the color image. After the transformation of the image, color image and depth image coincide. The depth value D of point P is extracted from the depth image, and 2D image coordinates of the central point P of parts are transformed to 3D coordinates (x_{P}, y_{P}, z_{P}) relative to Kinect with Kinect SDK toolkit, thus to have the location effect on parts.

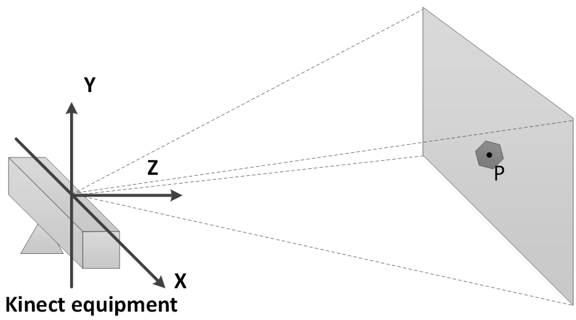

Fig. 8. Location schematic diagram

4. Extraction and identification of trajectories of disassembly tools

The trajectory extraction of hand-held disassembly tool includes the following main steps: the depth image of right arm and disassembly tools is extracted, to binarize the depth image; then the binary image is refined to acquire the position on the top of the refined image; the movement trajectory of disassembly tools is extracted, to carry out quantization coding of the trajectory; training is carried out and trajectory identification is realized with HMM algorithm.

4.1. Depth image of acquisition tools

The depth image contains an image with information on object surface distance in the scene. Compared with the color image, the depth image can reflect 3D feature of the object surface, which is not affected by external factors such as illumination. The depth image acquired with Kinect can recognize main parts of human body, and the trajectory of the hand-held disassembly tool is mainly studied in this paper, thus to identify the hand-held disassembly tool. The joint added with the tool is taken as the twenty-first joint as shown in Fig. 9. As the hand-held disassembly tool, in order to acquire more image information for facilitation of image processing, Kinect is used to extract the depth image of forearm and disassembly tool (with the screwdriver as an example) as shown in Fig. 10.

4.2. Image refinement

The feature of the acquired depth image of forearm and disassembly tool is not fine, which brings a great difficulty for the acquisition of top coordinates of disassembly tool, thus the image analysis method is introduced, to get a fine skeleton feature [8], and this is image refinement.

The image refinement is the refinement aiming at the binary image. For the binary image of 640×480, 1 represents the pixel region to be refined, and 0 represents the background. P_0, P_1, …, P_6, P_7 are defined as the eight neighborhood of point

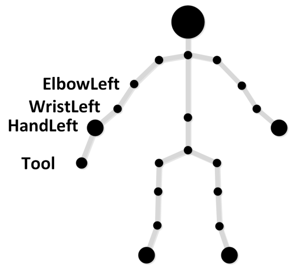

Fig. 9. Human body joints

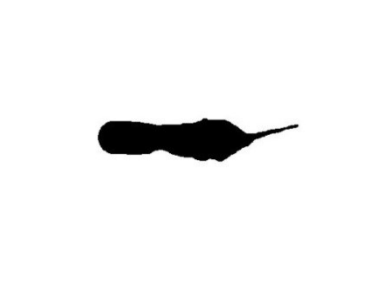

Fig. 10. Right forearm and disassembly tool

P. For the boundary point P, pixel values corresponding from P_0 to P_7 are judged, weights S of eight neighborhood points are calculated and checked with S value. If the number corresponds to 1, it can be judged that P is a deletable point

$$S = \sum_{k=0}^{7} P_k \times 2^k . \tag{2}$$

In the formula P_k indicates the pixel value of the k^{th} neighborhood point, k indicates the number of neighborhood point.

The binary image of forearm of human body and disassembly tool is refined with the above method, as shown in Fig. 11a.

4.3. Extraction of 3D coordinates on top of disassembly tools

For the refined image, the coordinates of points on the curve are acquired. With Kinect joint tracking, the coordinates of Point G of elbow joint can be extracted, to calculate the Euclidean distance d from Point G and other point on the curve, and Point p corresponding to the maximum Euclidean distance d is the top of disassembly tool, as shown in Fig. 11b.

Fig. 11. Forearm of human body: left–sketch map, right–finding top sketch map

4.4. Acquisition and Quantization coding of tool trajectory

The disassembly tool is held with hand for the movement (with the circle as an example). During the whole process of the movement, Kinect is used to extract a series of depth images to carry out operations of 4.2∼4.3 for each depth image, thus to acquire the position on the top of disassembly tool and extract the feature of the trajectory.

In order to facilitate the identification of the trajectory, the quantization coding of the trajectory is required. For the extracted trajectory, its depth value Z is removed, to facilitate the quantization coding. Fig. 12 gives the direction vector map of solving angle. $P\,(x_0,\, y_0)$ indicates the center of gravity of the trajectory curve. For any point $A\,(x_1,\, y_1)$ on the trajectory curve, the angle value θ of Point A is solved with the geometrical relation. A series of angle values are obtained on the trajectory curve, and 16-direction chain code is used to carry out angle quantization coding. Each point on the trajectory corresponds to an angle value, and the angle value is converted to the code value of 1 to 16 according to the correspondence, thus to identify the trajectory with the code value.

4.5. Trajectory identification

HMM (Hidden Markov Model) [9] is the probability model that describes the statistical characteristics of the stochastic process, which can describe the Markov process containing hidden unknown parameters [10]. First, hidden parameters of such process can be determined by observable parameters; then these parameters are used to carry out for the pattern recognition, et.

Five elements are used in HMM for the description, including two states and three state probability matrices.

(a) Number N in hidden state
 Hidden state $Q = \{1,\, 2,\, \cdots,\, N\}$.

(b) Number of observation symbols M
 Observation symbol $V = \{1,\, 2,\, \cdots,\, M\}$.

(c) Probability matrix in initial state P_i

$$\pi = \{\pi_i | i = 1,\, 2,\, \cdots,\, N\}\ . \tag{3}$$

In the formula π_i indicates the probability distribution in the initial state, and it meets $\sum_{i=1}^{N} \pi_i = 1$.

(d) Transfer probability matrix A in hidden state.

$$A = \{a_{ij} | i, j = 1, 2, \cdots, N\} . \tag{4}$$

In the formula: a_{ij} indicates the transfer probability at t moment in i state to $t+1$ moment in j state, and it meets $\sum_{i=1}^{N} a_{ij} = 1$.

(e) Probability matrix $\boldsymbol{B}$ in confused state

$$\boldsymbol{B} = \{b_j (k) | j = 1, 2, \cdots, N ; k = 1, 2, \cdots, M\} . \tag{5}$$

In the formula $b_j (k)$ indicates the probability observed in j state consistent with k, and it meets $\sum_{j=1}^{N} b_j (k) = 1$.

Before HMM is used for the trajectory identification, HMM parameters should be estimated. The forward-backward algorithm (Baum-Welch algorithm) can solve the parameter estimation of the hidden Markov model.

Given with observation sequence O and hidden Markov model λ, the probability variable at t moment in hidden state S_i is defined as $\gamma_t (i) = P (q_{\mathrm{t}} = S_i | O, \lambda)$, and this variable is converted to the following from forward variable and backward variable

$$\gamma_t (i) = \frac{\alpha_t (i) \beta_t (i)}{P (O|\lambda)} . \tag{6}$$

In them

$$P (O|\lambda) = \sum_{i=1}^{N} \alpha_t (i) \beta_t (i) . \tag{7}$$

Given with observation sequence O and hidden Markov modelλ, the probability variables in hidden state S_i at t moment and in hidden state S_j at $t+1$ moment are defined as

$$\gamma_t (i) = P (q_t = S_i, q_{t+1} = S_j | O, \lambda) . \tag{8}$$

In Formula (10), the variable is converted to the following from forward variable and backward variable:

$$\zeta_t (i, j) = \frac{\alpha_t (i) a_{ij} b_j (O_{t+1}) \beta_{t+1} (j)}{P (O|\lambda)} . \tag{9}$$

In them

$$P (O|\lambda) = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t (i) a_{ij} b_j (O_{t+1}) \beta_{t+1} (j) . \tag{10}$$

The quantization coding of multi-group angles is carried out to extract γ and

ζ by the substitution. These two variables are used to re-estimate HMM model parameters:

$$x_i = \gamma_1\left(i\right) , \tag{11}$$

$$a_{ij} = \frac{\sum\limits_{t=1}^{T-1} \zeta_t\left(i,j\right)}{\sum\limits_{t=1}^{T-1} \gamma_t\left(i\right)} , \tag{12}$$

$$b_j\left(k\right) = \frac{\sum\limits_{t=1}^{T} \gamma_t\left(j\right) \atop t=1.O_t=v_k}{\sum\limits_{t=1}^{T} \gamma_t\left(j\right)} . \tag{13}$$

$P\left(O|\lambda\right)$ reaches the maximum value through the trained model $\lambda = \left(\pi A, B\right)$, and its logarithm $\log P\left(O|\lambda\right)$ is usually taken. The likelihood calculation of test samples after quantization coding is carried out with Viterbi algorithm according to the trained model, and the trajectory model class with the maximum probability is taken as the class of such trajectory sequence, namely, identification result.

5. Experiment

5.1. Identification and location of typical parts

For the image shown in Fig. 13(a), the template shown in Fig. 13(b) is used for matching identification with the correlation coefficient method.

In the matching identification with the correlation coefficient method, first, when $R\left(x,\, y\right)$ is maximum. Then, the outline of the point with the maximum gray value is extracted. Finally, the position of the square nut in image coordinates is determined as shown in Fig. 14.

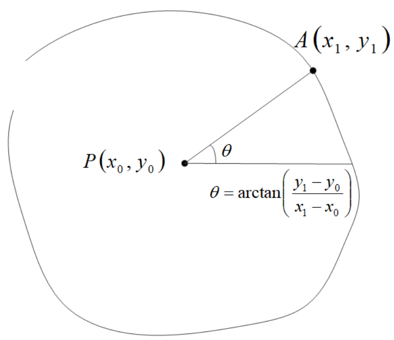

Fig. 12. Direction vector diagram

Fig. 13. Image matching: left–search graph, right–template

Fig. 14. Square nut identification

The image pixel coordinate of the center of the square nut identified with Kinect is (312, 107), with the unit of pixel, and the corresponding depth value is 1.035 m. Kinect SDK development toolkit can be used to convert image pixel coordinates to 3D coordinate, and the extracted spatial position coordinate relative to Kinect center is (−0.014, 0.241, 1.035), with the unit of m.

5.2. Movement trajectory identification of disassembly tools

Three trajectory types are designed in this experiment for the identification, including circular trajectory, vertical movement trajectory and horizontal movement trajectory, as shown in Fig. 15. By experimental analysis, threshold parameters are $T_1 = 60$, $T_2 = 300$, and $T^{'} = 20$. 80 groups of examples are respectively used in this study for training of each trajectory type. Table 1 shows the trajectory identification rate of HMM algorithm by the experimental analysis.

6. Conclusion

In this paper, the disassembly of waste mechanical products is carried out based on the autonomous robot; parts and the disassembly trajectory identification are

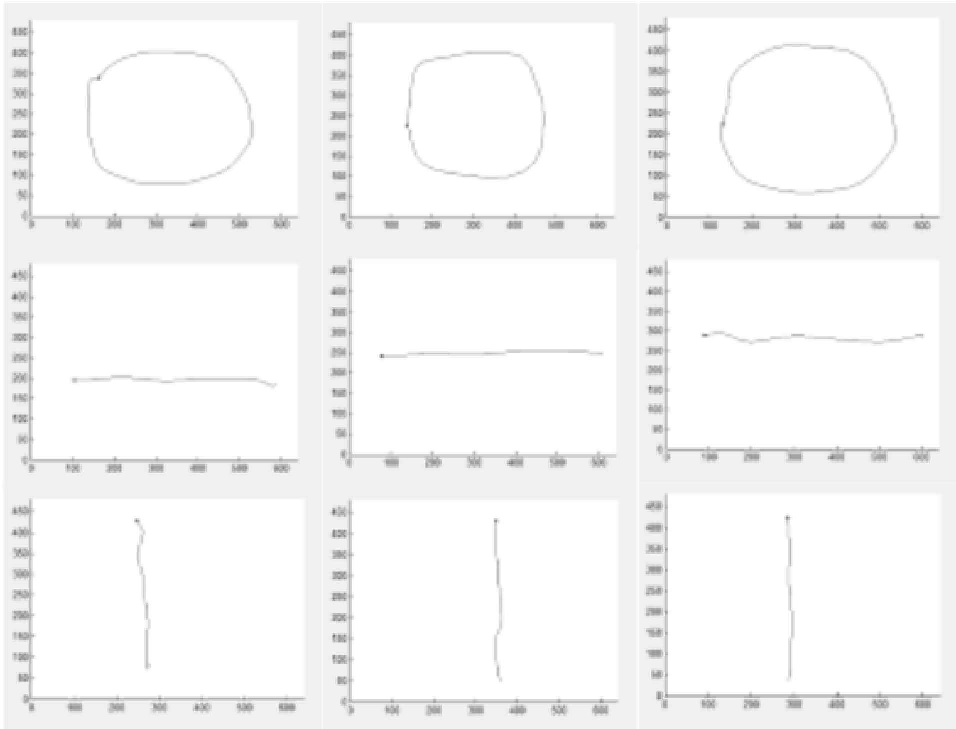

Fig. 15. Three trajectory diagrams

Table 1. Identification rate of HMM algorithm to three trajectories

Trajectory	Number of training samples	Number of test samples	Number of correct identification	Identification rate
Circular	80	50	41	82%
Transverse	80	50	44	88%
Vertical	80	50	44	88%

studied. The identification of parts and the spatial position is realized based on Kinect, and the depth image is used for the refinement to extract the trajectory feature of the disassembly tool, thus to realize the trajectory identification by training trajectory with the hidden Markov model.

Kinect has characteristics such as modularization, which is easy to redevelop, and a feasible solution is provided for autonomous robot oriented part identification and disassembly process by installing Kinect module with the existing robot system.

References

[1] J. Lu, V. E. Liong, X. Zhou, J. Zhou: *Learning compact binary face descriptor for face recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence *37* (2015), No. 10, 2041–2056.

[2] E. Smistad, T. L. Falch, M. Bozorgi, A. C. Elster, F. Lindseth: *Medical image segmentation on GPUs–A comprehensive review.* Medical Image Analysis *20* (2015), No. 1, 1–18.

[3] S. Cloix, G. Bologna, W. Weiss, T. Pun, D. Hasler: *Descending stairs detection with low-power sensors.* European Conference on Computer Vision, 6–12 September 2014, Zurich, Switzerland, Springer, ECCV 2014 Workshops, Part III, LNCS *8927* (2015), 658–672.

[4] M. Ye, Y. Zhang, R. Yang, D. Manocha: *3D reconstruction in the presence of glasses by acoustic and stereo fusion.* IEEE Conference on Computer Vision and Pattern Recognition, 7–12 June 2015, Boston, MA, USA, IEEE Computer Society (2015), 4885–4893.

[5] Y. Wu, M. Pei, M. Yang, J. Yuang, Y. Jia: *Robust discriminative tracking via landmark-based label propagationn.* IEEE Transactions on Image Processing *24* (2015), No. 5, 1510–1523.

[6] F. J. Lawin, P. E. Forssén, H. Ovrén: *Efficient multi-frequency phase unwrapping using kernel density estimation.* European Conference on Computer Vision, 11–14 October 2016, Amsterdam, The Netherlands, Springer, ECCV 2016, Part IV, LNCS *9908* (2016), 170–185.

[7] M. Dou, J. Taylor, H. Fuchs, A. Fitzgibbon, S. Izadi: *3D scanning deformable objects with a single RGBD sensor.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7–12 June 2015, Boston, MA, USA, IEEE Conference Publications (2015), 493–501.

[8] S. Zhang, X. Gao, N. Wang, J. Li: *Robust face sketch style synthesis.* IEEE Transactions on Image Processing *25* (2016), No. 1, 220–232.

[9] Y. Cao, Y. Li, S. Coleman, A. Belatreche, T. M. McGinnity: *Adaptive hidden Markov model with anomaly states for price manipulation detection.* IEEE Transactions on Neural Networks and Learning Systems *26* (2015), No. 2, 318–330.

[10] L. Rossi, J. Chakareski, P. Frossard, S. Colonnese: *A Poisson hidden Markov model for multiview video traffic.* IEEE/ACM Transactions on Networking *23* (2015), No. 2, 547–558.

Received April 30, 2017