KINECT 3D RECONSTRUCTION FOR QUANTIFICATION OF GRAPE BUNCHES VOLUME AND MASS

Francesco Marinello, Andrea Pezzuolo, Donato Cillis, Luigi Sartori
University of Padova, Italy
francesco.marinello@unipd.it

Abstract. In the last decade, three-dimensional reconstruction of plants has gained a noticeable importance, in particular for the possibility of collecting data correlated to biomass, leaves area, etc. Different sensing technologies are available for 3D imaging, as, for instance, laser scanning, or stereoscopic reconstruction: however, practical application is still limited by high costs, or high speed data processing demand. For the present work the depth sensing camera technology is implemented. Such technique has gained in the last years a wide diffusion in particular thanks to recent introduction in the video games industries. Specifically, a Microsoft Kinect™ RGB-depth was used. The Kinect sensor can be applied in fruitculture in order to allow fast reconstruction, allowing characterization of the canopy or estimation of the fruit volume. In this paper it is proposed specifically for vineyard applications, where strict national or regional regulations limit plant management and grape production. The work focuses in particular on the possibility of implementing such sensor in order to rapidly estimate the grape mass and allow foreseeing production before harvesting. To this end, measurements are carried out in the case of both black and white berries, with different mass grapes (ranging between 50 and 1000 g) and scanned at different distances with two orientations of the sensor.

Keywords: Kinect, vineyard, grapes, 3D reconstruction.

Introduction

Agriculture is constantly pushing research for development of technological solutions allowing automation of different tasks with the aim to help timeliness, increase production and reduce inputs. Such automation-oriented trend can benefit from constant miniaturization of sensors [1; 2], and is supported by researches coming from different fields, such as aerospace or military (as in the case of drones [3]), automotive (such as for electronics control and diagnostic [4]) or medicine (such as for biometrics analyses in livestock [5]). Automation is interesting for different agricultural operations, with the potential to improve soil preparation, watering, weed control, agro-chemical application or harvesting. In the latter case estimation of the quantities (as in the case of cereal crops), dimensions (as in the case of vegetable high-value crops), position and weight (as in the case of orchards) is ideally needed in order to allow machine optimization. With specific reference to dimensioning and positioning operations, different technologies are available. Since fast data collection and preservation of the product quality are the primary conditions, tactile instruments cannot be implemented, while a great range of possibilities is allowed by non contact sensors. Among these, multi-image triangulation, lidar technology, time of flight or fringe projection are only some of the many available possibilities.

The present paper implements a non-contact sensor based on depth-camera sensing. Nowadays, technology has reached a mature level and low-cost depth cameras developed for the videogames industry are available on the market. This work specifically takes advantage of the Microsoft Kinect™ RGB-depth camera [6; 7], which is already raising significant interest and growing attention in agriculture applications for its high quality and low cost [6-9].

In particular, in viticulture, the interest in determining the grape mass is related to the possibility of foreseeing the final yield and profitability [10; 11] thanks to implementation of specific simulation models [12; 13] and to the need to comply with the national or regional regulations, which limit plant management and grape production.

Materials and methods

The present work implements the low-cost Microsoft Kinect depth camera. Such kind of technology uses an infra-red (IR) projector emitting IR radiation onto the scene: radiation back reflected from the scene is captured by an IR depth sensor and allows a 3D reconstruction of the scene itself [6]. Representation of the Kinect sensor is proposed in Figure 1, with the working principle and reference axes (Figure 1a) and the main components (Figure 1b): the infrared laser emitter, the infrared camera and the RGB camera.

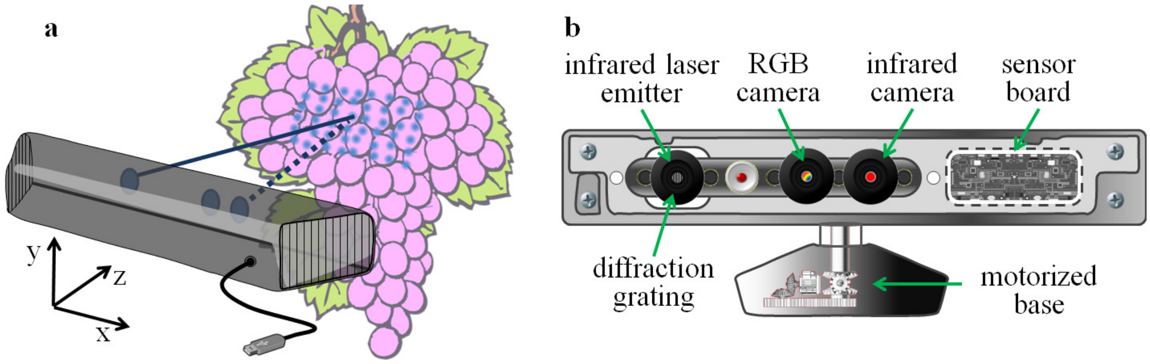

Fig. 1. **Kinect 3D sensing is based on a triangulation principle (a).; schematic view of the Kinect sensor with infrared (IR) projector, IR and RGB cameras localization (b)**

The Kinect depth sensing sensor has the merit to allow exploitation of static and dynamics characterizations of 3D surfaces. For such reason it can be considered as an eligible solution in viticulture applications, where three dimensional characterization of plants is still an open issue. Such need regards both the canopy and the fruits. With reference to the canopy, the possibility of knowing the canopy volume is important for properly dosing agrochemicals. On the other hand, accurately estimating the grape mass is of help in order to foresee the final yield and, as a consequence, properly schedule thinning or other operations. In this work a preliminary study is proposed, focused on the possibility of using the Kinect sensor for effective estimation of grape bunches mass. To this end, the sensor can be installed on a tractor or on a moving vehicle according to two possible configurations:

- bottom scanning, with the *x* and *y* axes lying on the horizontal plane, the *z* axis orthogonal to the soil and the grapes scanned from the bottom (Figure 2a),
- lateral scanning, with the x and z axes lying on the horizontal plane, the y axis orthogonal to the soil and the grapes scanned from the side (Figure 2b).

In the first case the main advantage is the accessibility to the large majority of grape bunches, while only in the second case the best view is achieved, allowing better volume reconstruction, even though the presence of leaves lying between the grapes and the sensor could negatively affect the measurement result. In this preliminary work, investigations were carried out encompassing three sets of experiments on:

- calibrated reference surfaces,
- grape clusters scanned from the side,
- grape clusters scanned from the bottom.

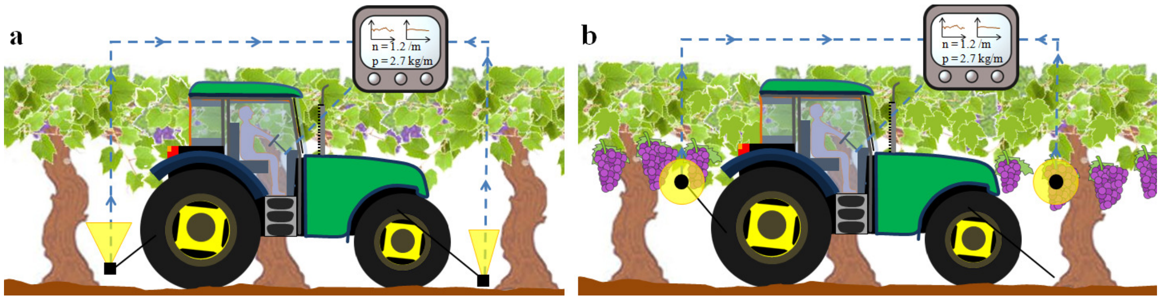

Fig. 2. **Two possible configurations with the three-dimensional sensor measuring from the bottom (a) or from the side (b)**

Calibrated reference surfaces

A set of tests was carried out in order to calibrate the scanner and allow extrapolation of quantitative data [14]. For the scope a set of measurements was done on 14 reference spherical elements. The spheres featured dimensions comparable with those of grape clusters, with radii ranging between 25 and 90 mm. Each reference surface was measured five times, with the Kinect sensor positioned at four different relative distances from the spheres, resembling the potential minimum and maximum scanning distance in field operations: 0.8, 1.2, 2.0 and 4.0 m.

Measurements on grape bunches from the side and from the bottom

The implemented sensor is based on an optical principle and can be sensitive to surfaces of different colors. For such reason, two different types of grapes were implemented: black Dan Ben Hannah grape and white Dauphine grape, both from South Africa. The main physical characteristics were measured for each variety (Table 1). Average diameter was estimated from the volume of 100 berries for each grape variety submerged in water in a graduated container; the same berries underwent mass measurement in order to calculate the actual density. The white grapes exhibit slightly larger berries, with an average diameter of about 22 mm, 1 mm larger than in the case of the black grape; the densities are quite similar, ranging in general between 1100 and 1180 $kg \cdot m^{-3}$. Apparent density was estimated by introducing 10 grape bunches for each variety in polymer film vacuum envelopes. The enveloped grapes were submerged in water in a graduated container and water displacement was compared to that coming from the same grapes submerged in water free from envelopes. The apparent densities are clearly lower than the actual densities, due to the effect of voids between neighbor berries (Table 1).

In both cases, eighteen different grape clusters were studied, with six different dimensions, ranging between 50 and 1000 g. The grape clusters were scanned from the side (see Figure 3a), with the Kinect sensor positioned at the same scanning distances considered during the calibration task (0.8, 1.2, 2.0 and 4.0 m). The same tests were repeated positioning the sensor on the bottom of the grape bunches, at a fixed distance of 0.8 m (see Figure 3b). The performed measurements were post-processed in order to compensate systematic calibration errors (as calculated by means of calibrated reference surfaces) and to quantify the grape volumes. Since the implemented sensor provides a three-dimensional convolution of surfaces and in particular cannot reproduce deep and hidden holes, the apparent density was implemented in order to estimate the grape mass.

Fig. 3. **Examples of 4 of the white grape bunches of different sizes seen from the side (a) and from the bottom (b)**

Table 1

Main data of grapes (standard deviations in brackets)

Variety	Color	Average berry mass [g]	Average berry diameter [mm]	Density [$kg \cdot m^{-3}$]	Apparent density [$kg \cdot m^{-3}$]
Dauphine	White	10.9 (2.4)	21.8 (2.0)	1158 (24)	789 (91)
Dan Ben Hannah	Black	9.0 (1.9)	20.8 (1.7)	1119 (22)	717 (87)

Results and discussion

Calibrated reference surfaces

Tests on the calibrated surfaces allowed determination of lateral and vertical resolutions. Such information is necessary in order to characterize the minimum volume detectable by the 3D camera. The lateral resolution in the XY plane (see Figure 1a), which determines the minimum detectable feature size, is determined by the point density and reduces as the distance of the sensor from the target surface increases. On the other hand, the depth resolution is determined by the computational capacity: the Kinect implements a 1/1024 approximation level, which gives different resolutions depending on the sample to the sensor distance. Reference values at different distances both for lateral and depth resolutions are shown in Figure 4a. It can be noticed how in general for measurements taken at distances of the sensor no larger than 2 m from the target surface provide resolutions no larger than 4 mm: such values are satisfactory for a proper three-dimensional reconstruction of grape clusters.

Tests on the calibrated surfaces allowed also verification of the procedures for volume determination, by comparing the measured values with the volumes calculated taking advantage of the reference diameter data. It was noticed that the Kinect sensor provides volumes which are systematically larger than the reference values (Figure 4b). This was ascribed to two factors: systematic distortion of the sensor and a convolution phenomenon of the scanning system, which causes a dilation of the scanned surfaces. Due to the systematic behavior, such deviation can be compensated while only deviations from linearity (i.e. non-linearities), which typically have a random behavior, cannot be compensated. Non-linearities tend to increase as the distance of the sensor from the target surface increases. In general, the best scanning condition corresponds to the lowest tested scanning distance (i.e. 0.8 m), with lateral and depth resolutions better than 1 mm, and average absolute deviations from linearity lower than 8 %.

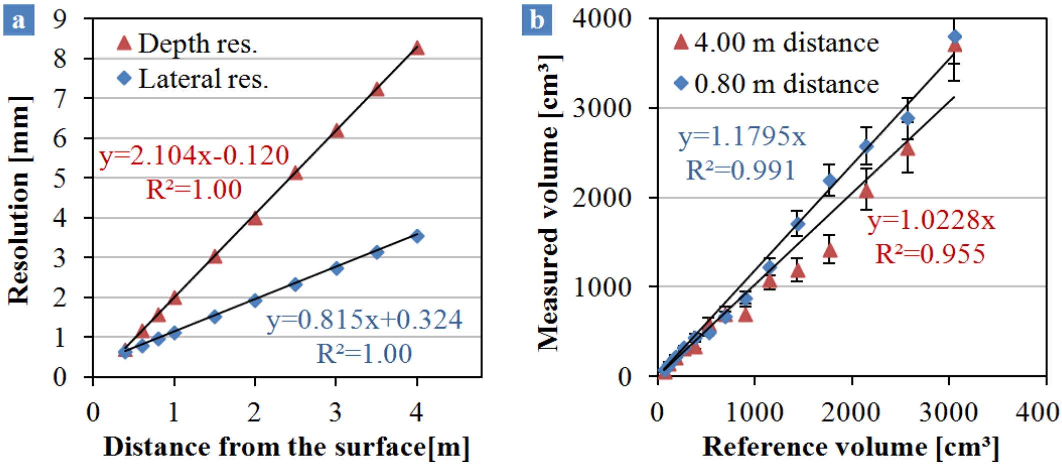

Fig. 4. **(a) Lateral and depth resolutions and (b) measured as a function of reference volumes at different relative distances of the sensor from the target surface. Error bars represent standard deviations of five repeated measurements**

Measurements on grape bunches from the side and from the bottom

A three-dimensional rendering of one of the reference samples, as revealed by the Kinect sensor, is proposed in Figure 5.

The proposed Figure 5 is merely qualitative, nevertheless, some considerations can be made. Grape cluster characterization from the bottom suffers from some limitations (Figure 5a): indeed, due to their elongated shape, grape clusters have a lateral surface which often exceeds the maximum detectable slope, causing a loss of signal, which limits the volume reconstruction. Additionally, irregularities are often detected also in lateral development, which introduce further uncertainty in volume quantification [11]. Such phenomenon causes a high level of uncertainty, which by now limits the practical application of such approach. This is demonstrated also by the graph of relative mass deviations from the reference (Figure 6a), where the values larger than 25-30 % are found: in particular, in the case of larger bunches (750-1000 g) three dimensional reconstruction was not possible. The problem is partially overcome in the case of lateral scanning, where relatively lower slopes of the scanned surface relative to the sensor XY plane are found. In this case relative mass deviations are comprised between 10 % and 15 %, both in the cases of white and black grapes (Figure 6a). The performance progressively worsens as the distance of the Kinect camera from the scanned surface increases (Figure 6b, 6c and 6d). In particular, relative deviations tend to increase in the case of smaller bunches: this is most probably due to the fact that at higher distances the influence of the resolution reduction plays a more critical role, negatively affecting mass estimation performance. However, it should be observed that in vineyards the row distances are typically lower than 2.5-3 m and a reasonable sensor positioning could be at a distance of 0.8-1 m from the plants, where it exploits its best performance. Additionally, higher uncertainties are detected in the case of smaller bunches, but on the other hand, their influence on the final yield is secondary if compared to larger bunches which conversely exhibit lower uncertainties and contribute for a 70-90 % on the total yield.

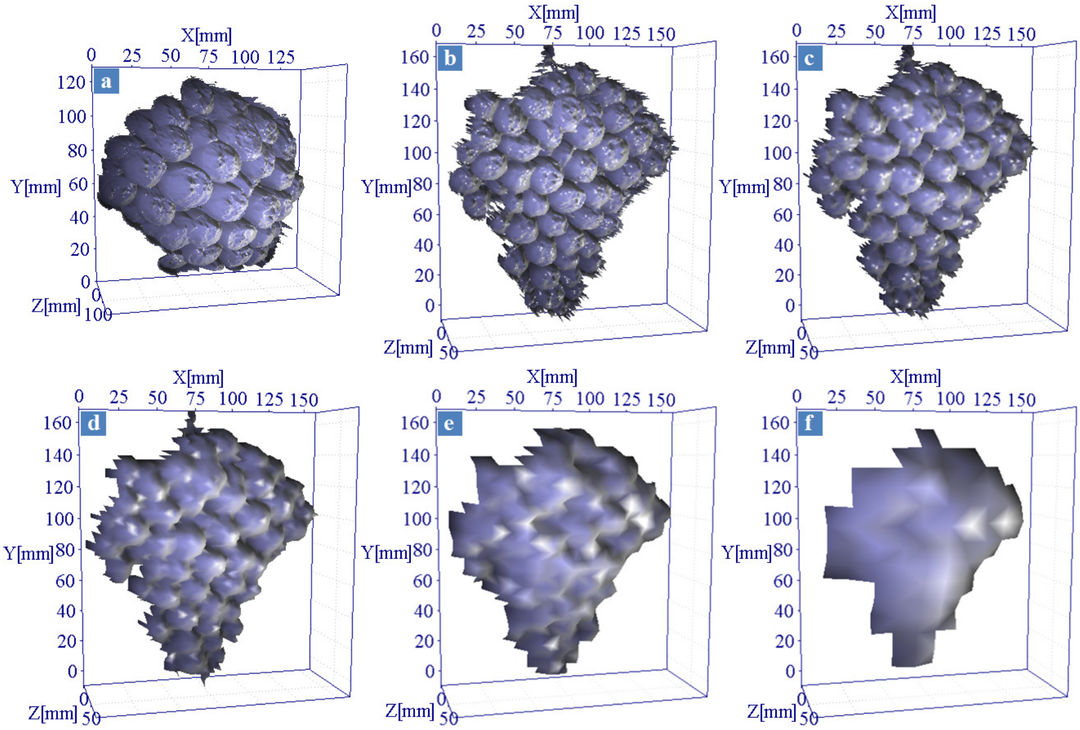

Fig. 5. **3D rendering of a grape cluster scanned from the bottom at a 0.8 m distance (a) and from the side at increasing distances: 0.8 m (b), 1.2 m (c), 2.0 m (d), 4.0 m (e) and 8.0 m (f)**

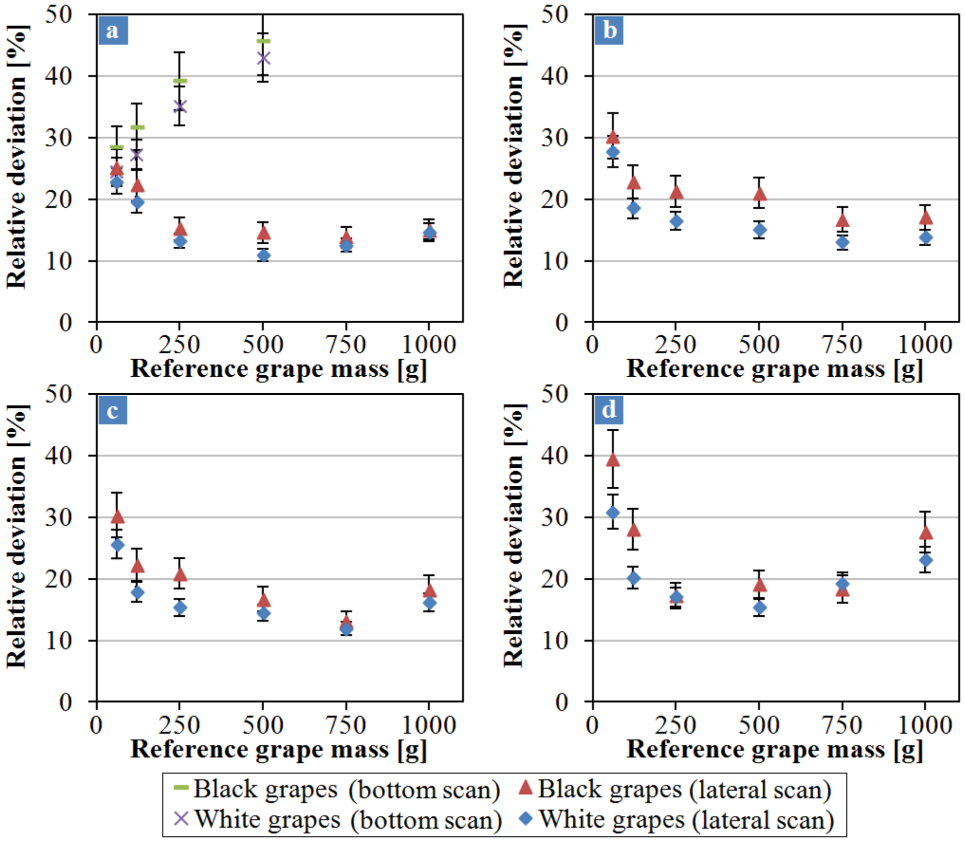

Fig. 6. **Average relative deviation of the estimated black and white grape bunches mass from the reference value, at different distances: (a) 0.8 m; (b) 1.2 m; (c) 2.0 m and (d) 4.0 m. Error bars represent standard deviations of three different bunches. BV and LV stand respectively for "bottom view" and "lateral view"**

Conclusions

1. The present work proposes implementation of the Microsoft Kinect™ RGB-depth camera as a cost effective solution for three-dimensional volume and mass characterization of grape clusters.
2. Tests on reference surfaces are reported to describe the sensor performance in terms of lateral and depth resolution and deviation from linearity.
3. A measurement campaign has been carried out on actual grapes featuring different colors (black and white varieties) and different dimensions (from 50 to 1000 g).
4. Extrapolated data highlight how deviations are very much influenced by the distance of the sensor from the grapes and by the dimensions of the bunches, and deviations as low as 10-15 % can be expected in the best scanning conditions.
5. Preliminary tests were carried out only on two different grapevine varieties and should be validated on different varieties with different bunches or berry shapes, however, the preliminary results pose promising perspectives.

Aknowledgement

The study is supported by a grant from the University of Padova (code CPDA143174), Italy. We also gratefully acknowledge the support of NVIDIA Corporation with donation of the Tesla K40 GPU used within the present research.

References

1. Marinello F., Schiavuta P., Cavalli R., Pezzuolo A., Carmignato S., Savio E., Critical factors in cantilever near-field scanning optical microscopy. IEEE Sensors Journal, vol. 14/9, 2014, pp. 3236-3244.
2. Pajares G., Peruzzi A., Gonzalez-de-Santos P. Sensors in Agriculture and Forestry. Sensor, vol. 13, 2013, pp. 12132–12139.
3. Marinello F., Pezzuolo A., Chiumenti A., Sartori L. Technical analysis of unmanned aerial vehicles (drones) for agricultural applications. Engineering for Rural Development, vol. 15, 2016.
4. Pitla S.K., Luck J.D., Werner J., Lin N., Shearer S. In-field fuel use and load states of agricultural field machinery. Computers and Electronics in Agriculture, vol. 121, 2016, pp. 290-300.
5. Park M.C., Ha O.K. Development of effective cattle health monitoring system based on biosensors. Advanced Science and Technology Letters, vol.117, 2015, pp.180-185.
6. Marinello F., Pezzuolo A., Gasparini F., Arvidsson J., Sartori L., Application of the Kinect sensor for dynamic soil surface characterization. Precision Agriculture, vol. 5, 2015, pp. 1-12.
7. Andújar D., Ribeiro A., Fernández-Quintanilla C., Dorado J. Using depth cameras to extract structural parameters to assess the growth state and yield of cauliflower crops Computers and Electronics in Agriculture, vol. 122, 2016, pp. 67-73.
8. Marinello F., Pezzuolo A., Simonetti A., Grigolato S., Boscaro D., Mologni O., Gasparini F., Cavalli R., Sartori L., Tractor cabin ergonomics analyses by means of Kinect motion capture technology. Contemporary Engineering Sciences, vol. 8/25-28, 2015, pp. 1339-1349.
9. Marinello F., Pezzuolo A., Cillis D., Gasparini F., Sartori L. Application of Kinect-Sensor for three-dimensional body measurements of cows, Proceedings of "7th European Conference on Precision Livestock Farming, ECPLF 2015", September 15-18, 2015, Milan, Italy, pp. 661-669.
10. Zacepins A., Stalidzans E., Karasha T. Profitability ranking of precision agriculture measurement system implementation, Engineering for Rural Development, vol. 12, pp. 164-169, 2013.
11. Nuske S., Achar S., Bates T., Narasimhan S., Singh S. Yield estimation in vineyards by visual grape detection. Proceedings of International conference on "Intelligent Robots and Systems 2011 IEEE/RSJ", September 25-30, 2011, San Francisco, USA, pp. 2352-2358.
12. Schöler F., Steinhage V. Automated 3D reconstruction of grape cluster architecture from sensor data for efficient phenotyping. Computers and Electronics in Agriculture, vol. 114, 2015, pp. 163-177.
13. Pezzuolo A., Basso B., Marinello F., Sartori L., Using SALUS model for medium and long term simulations of energy efficiency in different tillage systems. Applied Mathematical Sciences, vol. 8/129-132, 2014, pp. 6433-6445.
14. Carmignato S., Savio E. Traceable volume measurements using coordinate measuring systems, CIRP Annals – Manufacturing Technology, vol. 60/1, 2011, pp. 519-522.