DOI: 10.1109/IJCNN.2012.6252515

Low-cost Volume Estimation by Two-view Acquisitions: A Computational Intelligence Approach

Ruggero Donida Labati, *IEEE, Member*, Angelo Genovese, *IEEE, Member*,
Vincenzo Piuri, *IEEE, Fellow*, Fabio Scotti, *IEEE, Member*
Department of Information Technology
Università degli Studi di Milano
Milano, 20122, Italy.
{ruggero.donida, angelo.genovese, vincenzo.piuri, fabio.scotti}@unimi.it

***Abstract*—The estimation of the volume occupied by an object is an important task in the fields of granulometry, quality control, and archaeology. An accurate and well know technique for the volume measurement is based on the Archimedes' principle. However, in many applications it is not possible to use this technique and faster contact-less techniques based on image processing or laser scanning should be adopted.**

In this work, we propose a low-cost approach for the volume estimation of different kinds of objects by using a two-view vision approach. The method first computes a reduced three-dimensional model from a single couple of images, then extracts a series of features from the obtained model. Lastly, the features are processed using a computational intelligence approach, which is able to learn the relation between the features and the volume of the captured object, in order to estimate the volume independently of its position and angle, and without computing a full three-dimensional model.

Results show that the approach is feasible and can obtain an accurate volume estimation. Compared to the direct computation of the volume from the three-dimensional models, the approach is more accurate and also less dependent to the position and angle of the measured objects with respect to the cameras.

I. Introduction

The estimation of the volume occupied by an object can be an important and non-trivial task in many applications. In granulometry applications, for example, such estimation can be useful to determine the volume of the particles to be examined [1 – 4], while in industrial quality control scenarios it is necessary to guarantee that the manufactured object adheres to certain mass and volume requirements. In the food industry, the information regarding volume and weight can expose important defects [5, 6], or it is necessary in order to determine the amount of food intake [7 – 9].

In archaeology, the volume estimation is used to determine the information regarding the constituting material of an object [10, 11], which is useful in order to correctly date the manufact.

One of the most precise method to determine the volume of an object consists in using the Archimedes' Principle: by immersing an object into the water and then measuring the displacement in the fluid level, it is possible to obtain an accurate estimation of the volume of the object. However, the procedure is time consuming and scarcely automatable for the purposes of today's industries and applications. Moreover, in many applications this technique cannot be applied for different reasons: the fragility or porosity of the object (for example, ancient vessels or food), its inaccessibility, or the impossibility to move the object.

For these reasons, image processing methods for the volume estimation based on the reconstruction of a three-dimensional model have been studied in the literature. Methods such as shape-from-silhouette [2, 10 – 13] or based on stereoscopic vision [3, 4, 14 – 16] are among the most widespread. Image processing techniques for volume estimation based on stereology [1, 17] have also been proposed . Moreover, laser scan techniques have been studied [18, 19], especially for larger volumes.

However, it is not always possible to build a three-dimensional model with arbitrary accuracy, especially when dealing with low-cost hardware setups. Also, the position and angle of the object directly influences the quality of the reconstruction. For this reason, in this paper we propose an approach based on image processing and computational intelligence techniques. Our approach exploits the generalization capability of neural networks, which are able to learn the relation between features and volume, in order to reduce the effects of orientations and illuminations of the acquired objects on the reconstructed model, and also to avoid the need for a complete three-dimensional model. It is then possible to achieve a robust volume estimation with a simple setup based on two cameras. The method, outlined in Fig. 1, starts from the volume approximation obtained from the convex hull of the three-dimensional model, extracts a set of features from the model, and then processes them to obtain an estimation of the volume of the object.

The paper is structured as follows: in Section II a short review of the methods for three-dimensional reconstruction and for volume estimation is presented, while in Section III the proposed method is discussed. Section IV contains the experimental results, and in Section V conclusions and future

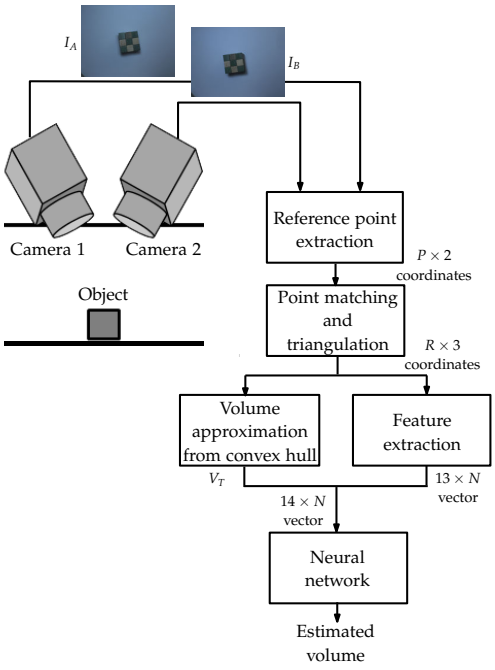

Fig. 1. Outline of the proposed method.

works are proposed.

II. Previous works

The techniques for the volume estimation of objects typically rely on the reconstruction of the three-dimensional corresponding model, because the correctness of the obtained measurement is strictly related to the quality of the estimated three-dimensional model. The majority of the techniques based on machine vision systems can be divided in shape-from-silhouette techniques, stereoscopic vision techniques, stereological approaches, feature-based methods, and single-view techniques.

Most of the techniques for the three-dimensional volume estimation are based on the shape-from-silhouette (SFS) reconstruction. The method is well-established, easy to implement, and relatively low-demanding in terms of computational complexity [12]. SFS techniques use the intersection of projection planes computed from multiple silhouettes, obtained by applying image segmentation algorithms. These three-dimensional reconstruction techniques can reconstruct only the shape of the object and not the details of the scene. Moreover, one of the major drawbacks is represented by the difficulty to reconstruct objects with concavities, since the concavities do not always influence the segmentation process [13]. SFS techniques are usually adopted when the objects that should be measured can be moved and complex setups can be employed. Examples of applications are the archaeological vessels [10, 11], or off-line analyses of irregularly-shaped objects [2].

Many methods for volume reconstruction are based on multiple-view techniques, which rely on the correspondences between pairs of images in order to determine the three-dimensional shape of the object. The images are generally captured at the same time and processed pairwise [14, 15]. Moreover, concave shapes can be represented as well. Multiple-view techniques are used when it is not possible to capture images of the object from every point of view, such as the case in which the object to be measured is lying on a plane. However, the most important problems of these techniques are related to the presence of occlusions and to the difficulty of searching correspondent points in different images. Examples of multiple-view systems for the volume estimation can be found in granulometry [3, 4], the food volume estimation [7], and in medical applications [16].

Stereological approaches are used when a certain number of bidimensional projections of the object on a plane are known. One of the uses of stereological techniques is in the determination of the three-dimensional information of particles placed on a conveyor belt, regardless of the their orientation [1].

Feature-based volume estimation methods are used when the position and orientation of the captured object is not known a-priori, while the information about the shape of the object that should be measured is available. The method proposed in [5] determines the volume of a pear placed on a conveyor belt by knowing a-priori the general shape of a pear, and extracting features related to shape, area and orientation of the measured pear. The method described in [6] describes the fruit as a composition of elementary elliptical frustums, and computes its volume as the sum of the elementary volumes.

Single-view volume estimation techniques are often used in the process of measuring the approximate volume of the food. The method described in [8] uses a single image of the food placed on a plate and considers the information about the known plate size to estimate the volume of the measured food. The method proposed in [9] uses an algorithm based on the creation of a virtual environment, in order to determine the volume of the food from a single image.

III. The proposed method

The proposed approach is designed to work with a two-view acquisition system, and it is able to capture a pair of images of an object placed on a flat surface. The approach is based on image processing and computational intelligence techniques.

The first step of the proposed volume estimation method consists in a three-dimensional reconstruction of the object, which uses a set of reference points extracted from the images, and then searches for correspondences using a matching algorithm. Then, a volume approximation is computed from the obtained model, and the features that describe the object shape are extracted. Lastly, computational intelligence techniques are used to refine the initial approximation.

The method is designed to work with low-cost acquisition setups and processing hardware, by using the computational intelligence techniques in order to learn the relation between the extracted features and the volume of the object. The extracted features, in fact, are almost invariant to the density of the three-dimensional model (as long as at least the main points are reconstructed) and to the position, angle, and illumination of the object. In this way, the method is capable

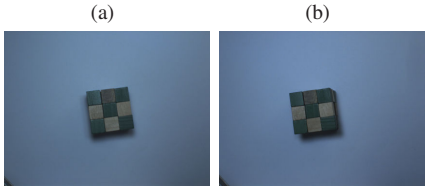

Fig. 2. A single two-view acquisition of a cube-shaped object: (a) image captured by the first camera; (b) image captured by the second camera.

of performing a volume estimation without computing a full three-dimensional reconstruction of the measured object, and also variations in the position and angle of the object can be compensated.

The proposed approach can be divided in the following steps:

1) camera calibration and acquisition;
2) extraction of the reference points;
3) point matching and triangulation;
4) volume approximation from the convex hull;
5) feature extraction.

A. Camera calibration and acquisition

The calibration of the cameras composing the acquisition setup is performed off-line. The calibration object used in the proposed method is a chessboard, captured using a two-view acquisition and processed using a corner detector. The intrinsic and extrinsic parameters of the stereoscopic system are then computed using the algorithms described in [20, 21]. The homography transformation matrix is computed from the extracted corner points using a DLT approach [14], while the fundamental matrix is computed from the extracted corner points using a RANSAC approach [22].

Each stereoscopic pair of images is then acquired by performing a single synchronized two-view capture. An example of the obtained results is shown in Fig. 2.

B. Extraction of the reference points

A small number of reference points is extracted from one of the two images in order to be subsequently matched and triangulated. The number of points is kept at a minimum (about 250), since for our purposes a full volumetric representation is not necessary.

A set of significative reference points can be extracted in two different ways. If the object surface is sufficiently variegated, the Harris corner detector is used to extract a sufficient number of points. In the case of more uniform surfaces, which cause the Harris method to perform poorly, a Canny edge detector is used to determine the reference points. A downsampling method is then used to extract a reduced set of points belonging to the edges.

In order to enhance the details, the first and the second image are preprocessed using a Sobel operator [23]:

$$\begin{aligned} I'_A &= I_A + (I_A * S) \ , \\ I'_B &= I_B + (I_B * S) \ , \end{aligned} \tag{1}$$

where S is the horizontal Sobel filter.

C. Point matching and triangulation

Several methods have been described in the literature for the matching of corresponding points in images captured using multiple view systems. In particular, the methods proposed in [24 – 26] deal with the aspects of matching points under different light conditions and with differences in the camera poses. However, the proposed setup presents small differences in the orientation of the cameras, and the illumination can be considered as uniform in the two images composing a two-view acquisition. In order to overcome these limitations, our approach uses a matching method based on the normalized cross-correlation.

In the search of the pairs of points with the highest correlation value, the method uses the information related to the homography and fundamental matrices computed during the calibration step , similarly to the methods presented in [27, 28]. Given a point x_A appertaining to the first image I'_A, the preliminary match of the corresponding point in the image I'_B is computed using the homography transformation, according to:

$$X'_B = H X_A \ , \tag{2}$$

where H is the 3×3 homography matrix, X_A is the point x_A expressed in homogeneous coordinates, and X'_B is the preliminary matching point estimation expressed in homogeneous coordinates:

$$X'_B = \begin{bmatrix} X \\ Y \\ W \end{bmatrix} \ . \tag{3}$$

The point X'_B is then converted in Cartesian coordinates, using the equation:

$$\mathbf{x}'_B = \begin{bmatrix} \frac{X}{W} \\ \frac{Y}{W} \end{bmatrix} \ . \tag{4}$$

The coordinates of the preliminary matching point are refined by considering the points adjacent to x'_B in the image I'_B. The considered points must appertain to a rectangular region centered in x'_B. A point is considered only if:

$$\begin{aligned} d_x(x^i_B, x'_B) < \Delta_x \\ d_y(x^i_B, x'_B) < \Delta_y \ , \end{aligned} \tag{5}$$

where x^i_B is the i-th adjacent point, d_x and d_y are the distances along the x and y axes, Δ_x and Δ_y are two empirically estimated values.

For each of these points x^i_B, the distance from the epipolar line corresponding to x_A is computed using the equation [14]:

$$d^i_{ep} = \frac{(X^i_B)^T F X_A}{\sqrt{(l_1)^2 + (l_2)^2}} \ , \tag{6}$$

where d^i_{ep} is the epipolar distance relative to the i-th adjacent point, X^i_B is the i-th adjacent point expressed in homogeneous coordinates, F is the fundamental matrix, and l_1, l_2 are the first two components of the epipolar line l, computed as:

$$l = F X_A \ . \tag{7}$$

The epipolar distance must be inferior to an empirically estimated threshold t_{ep}:

$$d^i_{ep} < t_{ep} \ . \tag{8}$$

Another check for the consistency of the candidate match points is performed by comparing the images obtained by applying the Canny edge detector on I_A and I_B. The binary values of the edge images at the positions x_A and x^i_B must be equal:

$$C_A(x_A) = C_B(x^i_B) \ , \tag{9}$$

where C_A, C_B are the images obtained by applying the Canny edge detector on I_A and I_B.

If the distances $d_x(x^i_B, x'_B), d_y(x^i_B, x'_B)$ of the point to the candidate point x'_B, and the epipolar distance d^i_{ep} are less than a fixed threshold, and the corresponding edge values are equal, the point x^i_B is included in the set of valid adjacent points:

$$x^i_B \in V_B \text{ if } \begin{cases} d^i_{ep} < t_{ep} \ , \\ d_x(x^i_B, x'_B) < \Delta_x \ , \\ d_y(x^i_B, x'_B) < \Delta_y \ , \\ C_A(x_A) = C_B(x^i_B) \end{cases} , \tag{10}$$

where V_B is the set of valid points adjacent to x'_B.

The cross-correlation of a $l \times l$ squared window centered in x_A and the $l \times l$ squared windows centered in each of the valid adjacent points belonging to V_B is then computed. The matching point is defined as the point with the maximum cross-correlation coefficient, computed using the formula:

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \ , \\ 1 < m < l, \ \ 1 < n < l \ , \tag{11}$$

where A and B are the two windows of size $l \times l$. The cross-correlation coefficient is chosen as the maximum of the coefficients computed on the Y, R, G, B channels separately.

To avoid high correlation values in presence of uniform surfaces, which would result in false matches, the local variance of a correlation window must be greater than a fixed threshold:

$$\sigma^2(A) > t_s \ , \tag{12}$$

where $\sigma^2(A)$ is the local variance of the window A and t_s is the fixed threshold.

The outliers are then removed by considering the global mean and standard deviation of the Euclidean distances between the points of the matched pairs. A pair of matched points p_i, composed by the points x_A of the first image and x_B of the second image, is considered as valid only if:

$$\bar{D}_S - 2\sigma_{DS} < d(x_A, x_B) < \bar{D}_S + 2\sigma_{DS} \ , \tag{13}$$

where $d(x_A, x_B)$ is the Euclidean distance between the points x_A and x_B, and $\bar{D}_S$ and σ_{DS} are the mean and the standard deviation of the Euclidean distances between the points of the matched pairs.

The two-dimensional coordinates of the matched pairs of points are refined with a rectification procedure that considers the calibration data. The z coordinate for each matched pair is then computed using a triangulation equation:

$$z = \frac{fT}{d(x_A, x_B)} \ , \tag{14}$$

where f is the focal length of the two cameras, T is the baseline distance between the two cameras, x_A and x_B are the two matched points, and d represents the Euclidean distance.

The homography matrix H is then used to recover the three-dimensional locations of the points belonging to the plane on which the object is placed. The point cloud is obtained by extracting a set of points on the image I'_A and computing the corresponding points on the image I'_B, using the equation 2. The points are then converted in Cartesian coordinates using the equation 4, and triangulated using the equation 14.

Some examples of reconstructed point clouds are shown in Fig. 3.

D. Volume approximation from the convex hull

A first volume approximation is computed from the convex hull of the reconstructed model, then the information is integrated with additional features in order to refine the estimation.

First, the three-dimensional Delaunay triangulation of the convex hull of the three-dimensional point cloud is computed. A three-dimensional Delaunay triangulation consists in the computation of a set of tetrahedrons, in which the vertices are the coordinates of the three-dimensional points P_i, with the constraint that no point P_i is inside the circumsphere of any tetrahedron.

The volume of the internal region delimited by the triangulation is obtained by summing the volume of each tetrahedron:

$$V_T = \sum_{i=1}^{N_T} V_i \ , \tag{15}$$

where V_T is the volume approximation of the object, N_T is the number of tetrahedrons and V_i is the volume of the i-th tetrahedron, computed as:

$$V_i = \frac{|(a-d) \cdot ((b-d) \times (c-d))|}{6} \ , \tag{16}$$

where a, b, c, d are the three-dimensional coordinates of the i-th tetrahedron.

E. Feature extraction

The features are extracted from the computed point cloud, and designed in order to create a fast computable description set, which can be used by the computational intelligence techniques to perform an accurate volume estimation. The extracted features are based on the computation of the three-dimensional bounding ellipsoid, a sphere fitting algorithm, and a plane interpolation technique.

1) Three-dimensional bounding ellipsoid: The three-dimensional bounding ellipsoid is computed by using a minimization problem [29]:

$$\min \left(\log \left(\det A \right) \right) \ , \tag{17}$$

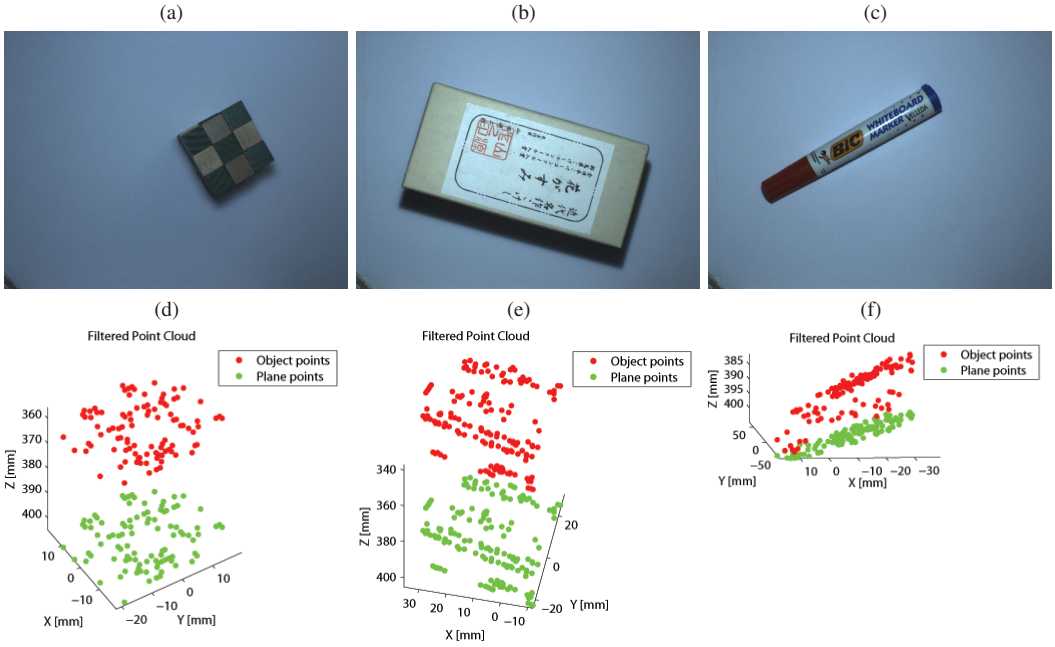

Fig. 3. Examples of captured object and the relative reconstructed point clouds: (a),(d) a first parallelepiped-shaped object and its reconstructed point cloud; (b),(e) a second parallelepiped-shaped object and its reconstructed point cloud; (c),(f) a cylinder-shaped object and its reconstructed point cloud.

where A is the 3×3 matrix of the ellipse equation, expressed in the form:

$$(x - c_e)^T A (x - c_e) = 1 \quad , \tag{18}$$

where c_e is the vector containing the coordinates of the center of the ellipse. The minimization problem is subject to the following constraints:

$$(P_i - c_e)^T A (P_i - c_e) \leqslant 1 \quad , \tag{19}$$

where P_i is the i-th point in the point cloud.

The singular value decomposition of the matrix A is computed:

$$[UQV] = \text{svd}(A) \quad , \tag{20}$$

then the radii r_1, r_2, r_3 of the bounding ellipsoid are computed using the following equations:

$$r_1 = \frac{1}{\sqrt{Q_{(1,1)}}} ; r_2 = \frac{1}{\sqrt{Q_{(2,2)}}} ; r_3 = \frac{1}{\sqrt{Q_{(3,3)}}} \quad . \tag{21}$$

2) Sphere fitting: The comparison between the point cloud and the relative fitted sphere can provide an estimation of the level of irregularity of the surface.

A sphere is fitted to the three-dimensional point cloud by using a least squares approach to solve the overdetermined system of normal equations:

$$x_i^2 + y_i^2 + z_i^2 + ax + by + cz + d = 0 \quad , \tag{22}$$

where x_i, y_i, z_i are the coordinates of the i-th point of the point cloud. The radius r_s of the sphere is computed as:

$$r_s = \sqrt{\frac{a^2 + b^2 + c^2}{4 - d}} \quad , \tag{23}$$

and the center c_s is computed as:

$$c_s = \begin{bmatrix} -\frac{a}{2} & -\frac{b}{2} & -\frac{c}{2} \end{bmatrix} \quad . \tag{24}$$

For each point P_i of the point cloud, the absolute difference d_i between the distance of the point from the center c_s and the radius r_s is computed:

$$d_i = | \, d(P_i, c_s) - r_s \, | \quad . \tag{25}$$

The minimum min_d, maximum max_d, mean $mean_d$ and standard deviation std_d of the obtained difference values are then computed.

3) Plane interpolation: In order to model the main inclination of the three-dimensional point cloud, thus correcting the volume estimation, a plane is fitted through the three-dimensional point cloud, using a first order polynomial interpolation. The resulting plane is in the form:

$$f(x, y) = p_{00} + p_{10} x + p_{01} y \quad . \tag{26}$$

4) Feature set: For each reconstructed point cloud, computed from a two-view capture of an object, 14 features are extracted:

- $F(1)$: the volume approximation V_T computed from the convex hull of the point cloud;
- $F(2-4)$: the lengths of the three main radii r_1, r_2, r_3 of the three-dimensional bounding ellipsoid;
- $F(5)$: the ratio of the length of the first radius to the length of the second radius (r_1/r_2);
- $F(6)$: the ratio of the length of the first radius to the length of the third radius (r_1/r_3);
- $F(7)$: the ratio of the length of the second radius to the length of the third radius (r_2/r_3);

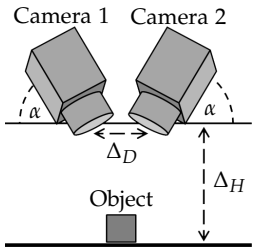

Fig. 4. Visual representation of the proposed acquisition setup.

- $F(8)$: the minimum value min_d of the differences between the radius of the sphere fitted to the data and the distance of each point from the center of the sphere;
- $F(9)$: the maximum value max_d of the differences between the radius of the sphere fitted to the data and the distance of each point from the center of the sphere;
- $F(10)$: the mean value $mean_d$ of the differences between the radius of the sphere fitted to the data and the distance of each point from the center of the sphere;
- $F(11)$: the standard deviation value std_d of the differences between the radius of the sphere fitted to the data and the distance of each point from the center of the sphere;
- $F(12-14)$: the coefficients p_{00}, p_{10}, p_{01} of the interpolating plane.

IV. Experimental results

The used acquisition setup is composed by two Sony XCD-SX90CR CCD color cameras synchronized by using a trigger mechanism. Both the cameras are angled of $\alpha = 85°$ with respect to the horizontal support, separated by a baseline of $\Delta_D = 80$ mm (the measure is taken considering the centers of the CCDs). The distance from the cameras to the flat surface is $\Delta_H = 395$ mm. A uniform illumination was used. A visual representation of the setup is shown in Fig. 4.

The calibration image set is composed by 15 different pairs of chessboard images. The calibration chessboard is composed by 12×9 squares of 10.5×10.5 mm. Considering these images, we estimated a reconstruction error of the chessboards in the three-dimensional space equal to 0.019 mm. This error is computed by triangulating the two-dimensional coordinates of the chessboard corners in the two views and considering the plane interpolating the three-dimensional corner positions. The standard deviation of the Euclidean distance between the triangulated corners and the interpolated plane is assumed as the error measure, as described in [30].

In order to prove the validity of the proposed method, we collected a dataset of 52 objects of various shape typologies (Fig. 5). We manually classified the objects into four categories according to their shape: parallelepiped-shaped objects (Fig. 5a-g), cylinder-shaped objects (Fig. 5h-m), sphere-shaped objects (Fig. 5n-s), and mixed-shape objects (Fig. 5t-v). Each object was captured 30 times in different positions and at different angles, for a total of 1560 stereoscopic acquisitions. A summary of the used dataset, with the volume range for each category, is shown in Table I.

TABLE I
Summary of the dataset

Object category	Volume (mm^3) Min	Max	No. of objects	Acquisitions each	Total no. of acquisitions
Parallelepiped-shaped	9,818	575,400	16	30	480
Cylinder-shaped	2,640	487,939	18	30	540
Sphere-shaped	16,619	197,350	13	30	390
Mixed-shaped	7,921	278,737	5	30	150

The volume of the parallelepiped-shaped objects was computed by measuring and then multiplying the dimensions of the three-edges. The volume of the cylinder-shaped objects, the sphere-shaped objects and the mixed-shaped objects was computed using the technique based on the Archimedes' Principle.

For each pair of images relative to a two-view acquisition, we reconstructed the three-dimensional point cloud and extracted the features using the described method. The features are extracted from each of the 30 acquisitions of every object, for a total of 1560 samples with 14 features each. We used a neural approach to determine the volume estimation and compared the results with the real volume of the considered object, with the aim to correct and reduce the effect of rotations and positions of the object. The generalization capability of the neural networks can, in fact, drastically reduce the effects of these problems.

The results depicted in the paper are computed using a N-fold cross-validation technique [31] with $N = 10$. The cross-validation was performed on each object category separately, using all the samples of each category. In this manner, we tested the ability of the neural network to generalize a particular type of objects (for example, parallelepiped, cylinder, sphere, or mixed) and then map the feature set of the particular sample into the corresponding volume.

The proposed approach is based on a Feed Forward Neural Network with one input layer, one hidden layer and one output layer. The input layer is composed by 14 nodes, while we tested different number of tan-sigmoidal nodes in the hidden layer: 1, 3, 5, 10, 15, 20, 25. The output layer is composed by one linear node. We used neural networks with a single hidden layer since they can be considered as universal approximators. The neural networks are trained with a Levenberg-Marquardt back-propagation algorithm, using at most 150 epochs. For each sample, the relative error e is computed as the absolute value of the difference between the real and the estimated volume:

$$e = \frac{|v_n - v_r|}{v_r} \quad , \tag{27}$$

where v_n is the output of the neural network and v_r is the real volume of the object.

We compared the results of the proposed method with the results obtained by directly approximating the volume from the convex hull of each point cloud, using the equation 15 and the equation 16. A summary of the results is depicted in Table II, showing the error values for the different configurations proposed.

Table II shows that the best configuration of the neural networks achieved a volume estimation of the parallelepiped-

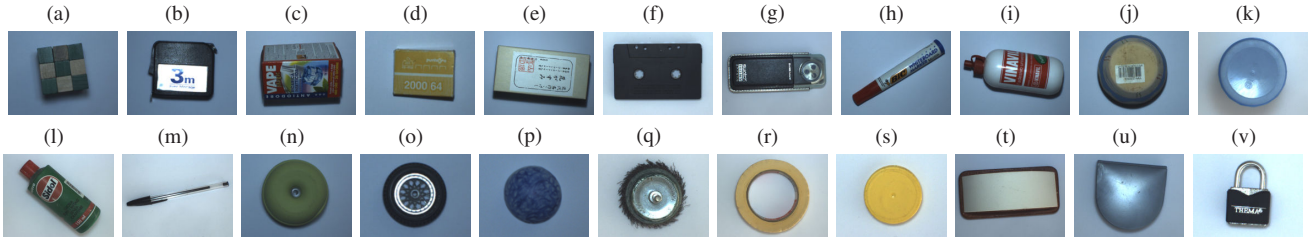

Fig. 5. Examples of captured objects belonging to different shape typologies: (a-g) parallelepiped-shaped objects; (h-m) cylinder-shaped objects; (n-s) sphere-shaped objects; (t-v) mixed-shaped objects.

TABLE II
RESULTS OF THE VOLUME ESTIMATION USING THE PROPOSED METHOD

Method	Object category	Relative error e Mean	Std
NN-1	Parallelepiped-shaped	0.048	0.051
	Cylinder-shaped	0.023	0.022
	Sphere-shaped	0.060	0.048
	Mixed-shaped	0.150	0.059
NN-3	Parallelepiped-shaped	0.023	0.032
	Cylinder-shaped	0.015	0.016
	Sphere-shaped	0.029	0.035
	Mixed-shaped	0.024	0.037
NN-5	Parallelepiped-shaped	0.029	0.036
	Cylinder-shaped	0.010	0.010
	Sphere-shaped	0.027	0.028
	Mixed-shaped	0.018	0.034
NN-10	Parallelepiped-shaped	0.019	0.022
	Cylinder-shaped	0.008	0.008
	Sphere-shaped	0.025	0.027
	Mixed-shaped	0.004	0.020
NN-15	Parallelepiped-shaped	0.014	0.019
	Cylinder-shaped	0.006	0.010
	Sphere-shaped	0.018	0.021
	Mixed-shaped	0.013	0.028
NN-20	Parallelepiped-shaped	0.019	0.032
	Cylinder-shaped	0.005	0.006
	Sphere-shaped	0.020	0.027
	Mixed-shaped	0.006	0.018
NN-25	Parallelepiped-shaped	0.026	0.041
	Cylinder-shaped	0.009	0.010
	Sphere-shaped	0.030	0.046
	Mixed-shaped	0.014	0.017
Direct volume approximation	Parallelepiped-shaped	0.500	0.567
	Cylinder-shaped	0.278	0.274
	Sphere-shaped	0.260	0.190
	Mixed-shaped	0.193	0.161

shaped objects with a mean error equal to 1.4 %, and the volume of the cylinder-shaped objects, the sphere-shaped objects and the mixed-shaped objects with a mean error less than 1 %. It is possible to observe that the error obtained by directly approximating the volume from the convex hull is much larger than the error obtained by using the neural approach. This is due to the fact that the volume computed directly from the three-dimensional reconstruction is strictly related to the position and the angle of the measured object with respect to the cameras. The proposed feature extraction process and the used computational intelligence techniques can achieve instead a more accurate and view-independent volume estimation.

For each object category, we used the best configuration of the neural networks and used it to compute the obtained volume estimation error for every object. The obtained results are depicted in Table III, showing that the proposed method obtained a remarkable accuracy for all the considered objects. The bigger estimation errors are related to objects with the most uniform surfaces, like the ink dryer, which are more difficult to reconstruct.

TABLE III
RESULTS OF THE VOLUME NEURAL ESTIMATION FOR EACH OBJECT

Object category	Object	Real vol. (mm^3)	Est. vol (mm^3) (*) Mean	Std
Parallelepiped-shaped	Cube	85,184	82,545	5,623
	Tape-line	69,231	64,305	28,206
	Box N.1	192,183	192,755	27,286
	Box N.2	24,805	23,487	15,470
	Box N.3	523,380	517,556	20,029
	Box N.4	134,400	132,826	6,252
	Box N.5	101,430	102,633	8,451
	Tape	49,600	43,665	9,240
	CD holder	173,430	169,174	19,030
	Book 1	575,400	577,402	9,683
	Book 2	234,520	239,784	28,250
	Memory case	42,248	39,395	7,638
	Coaster	43,264	46,947	13,834
	Watch case	153,149	150,557	11,293
	Keychain	9,818	7,601	9,390
	Walkman	267,814	267,916	18,274
Cylinder-shaped	Felt-tip pen N.1	30,769	29,995	3,604
	Felt-tip pen N.2	20,835	19,880	3,782
	Glue container	146,150	145,488	4,769
	Vase	101,340	101,166	3,152
	Glass	427,055	426,855	5,854
	Screwdriver	2,640	3,120	6,558
	Tin can	487,939	486,319	1,487
	Pen N.1	10,562	9,631	3,638
	Pen N.2	5,281	6,547	2,508
	Pen N.3	10,562	11,699	2,951
	Pen N.4	15,843	15,628	4,768
	Pen N.5	21,124	19,585	3,698
	Brush	2,640	3,669	2,037
	Battery	8,305	8,534	2,882
	Brown tape	298,024	297,308	5,209
	Small bottle	242,926	242,971	3,663
	Filler bottle N.1	212,371	212,741	5,622
	Filler bottle N.2	292,011	292,915	3,219
Sphere-shaped	Toy wheel	16,619	17,157	1,816
	Headphones holder N.1	84,615	84,564	5,375
	Headphones holder N.2	79,215	79,330	3,720
	Ping-pong ball (half)	16,755	18,186	6,626
	Round tape-line	159,241	159,386	13,116
	Paper tape	197,350	193,539	9804
	Small tape	38,813	38,486	4,173
	Sandpaper brush	66,366	63,537	9,743
	Metal brush	86,205	85,657	4,006
	ADSL plug	68,653	66,091	9,027
	Top N.1	23,524	22,653	5,003
	Top N.2	21,598	21,995	3,170
	Yo-yo	76,340	72,676	9,366
Mixed-shaped	Ink dryer	278,737	274,811	21,497
	Lock N.1	7,921	7,928	12
	Lock N.2	10,562	10,496	357
	Cash holder	105,620	105,622	6
	Sunglasses holder	156,180	156,296	679

(*) Processed 30 acquisitions of each object with different angles and positions.

V. CONCLUSION

In this paper we proposed a low-cost approach for the volume estimation of objects based on a single two-view acquisition. The method is designed in order to achieve an accurate and view-independent volume estimation, without the need to compute a full three-dimensional reconstruction of the object, using complex setups or time-consuming algorithms.

The method uses image processing and computational intelligence techniques, and it is based on a fast three-dimensional reconstruction step and in a feature extraction process. The extracted features, along with a first approximation of the volume computed directly from the point cloud, are used by a

neural network in order to estimate the volume of the object.

Each neural network has been trained to correct the initial convex hull estimation of the object for four general shape types, in particular parallelepiped, cylinder, sphere and mixed-shaped objects. Within the correct shape type, experiments showed that the trained neural network is capable to effectively correct the volume estimation for different objects. The neural approach permitted then an accurate volume estimation of the objects which is invariant to orientation, position, and illumination, and using a reduced three-dimensional reconstruction.

In order to test the validity of the proposed method, we captured different objects and classified them in separate categories according to their shape. Then, we performed the tests on the extracted features using different configurations of the neural networks. We performed the tests on each category separately. We compared the results with the ones obtained by computing the volume approximation directly from the three-dimensional reconstruction. The results show that the proposed method achieves a better accuracy in estimating the volume of the objects with respect to the direct volume approximation from the point cloud, proving that the method is feasible.

However, in the case of complex-shaped objects or major occlusions, the proposed approach produce a less accurate volume estimation since the extracted feature set is not sufficient to describe the complexity of the surface, in order to achieve a robust and accurate neural correction of the convex hull method. Future works will focus on achieving an accurate volume estimation with more complex objects.

REFERENCES

[1] B. Presles, J. Debayle, A. Cameirao, G. Fevotte, and J. Pinoli, "Volume estimation of 3D particles with known convex shapes from its projected areas," in *2010 2nd International Conference on Image Processing Theory Tools and Applications (IPTA)*, July 2010, pp. 399–404.

[2] E. Castillo-Castaneda and C. Turchiuli, "Volume estimation of small particles using three-dimensional reconstruction from multiple views," in *Proceedings of the 3rd international conference on Image and Signal Processing*, ser. ICISP. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 218–225.

[3] C. Wang and Y. Han, "Design of dynamic volume measure system based on binocular vision," in *2010 International Conference on Computer Application and System Modeling (ICCASM)*, vol. 10, October 2010, pp. 289–293.

[4] M. Kempkes, T. Vetter, and M. Mazzotti, "Measurement of 3D particle size distributions by stereoscopic imaging," *Chemical Engineering Science*, vol. 65, no. 4, pp. 1362–1373, February 2010.

[5] K. Forbes and G. Tattersfield, "Estimating fruit volume from digital images," in *1999 IEEE Africon*, vol. 1, 1999, pp. 107–112.

[6] M. Omid, M. Khojastehnazhand, and A. Tabatabaeefar, "Estimating volume and mass of citrus fruits by image processing technique," *Journal of Food Engineering*, vol. 100, no. 2, pp. 315–321, 2010.

[7] M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney, "Recognition and volume estimation of food intake using a mobile device," in *2009 Workshop on Applications of Computer Vision (WACV)*, December 2009, pp. 1–8.

[8] Y. Yang, Y. Yue, Z. Wei, J. Robert, W. Jia, and M. Sun, "Food volume calculation in different imaging scenarios," in *2011 IEEE 37th Annual Northeast Bioengineering Conference (NEBEC)*, April 2011, pp. 1–2.

[9] Z. Zhang, Y. Yang, Y. Yue, J. Fernstrom, W. Jia, and M. Sun, "Food volume estimation from a single image using virtual reality technology," in *2011 IEEE 37th Annual Northeast Bioengineering Conference (NEBEC)*, April 2011, pp. 1–2.

[10] S. Tosovic and R. Sablatnig, "Volume estimation for objects with concavities," in *Proc. of the 6th Computer Vision Winter Workshop 2001*, B. Likar, Ed. Bled, Slovenia: Slovenian Pattern Recognition Society, February 2001, pp. 49–59.

[11] R. Sablatnig, S. Tosovic, and M. Kampel, "Combining shape from silhouette and shape from structured light for volume estimation of archaeological vessels," in *Proceedings of the 16th International Conference on Pattern Recognition (ICPR)*, vol. 1, 2002, pp. 364–367.

[12] C. Nitschke, *3D Reconstruction - Real-time Volumetric Scene Reconstruction from Multiple Views*. VDM Verlag Dr. Muller, April 2007.

[13] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 150–162, February 1994.

[14] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[15] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, June 2006, pp. 519–528.

[16] K. Pirker, M. Rüther, H. Bischof, F. Skrabal, and G. Pichler, "Human body volume estimation in a clinical environment," in *AAPR/OAGM: Challenges in the Biosciences: Image Analysis and Pattern Recognition Aspects, Stainz Austria*. Austrian Computer Society, 2009.

[17] J. C. Russ and R. T. DeHoff, *Practical Stereology*, 2nd ed. New York: Kluwer Academic / Plenum Publishers, 2000.

[18] W. Lulu and W. Bin, "Research on estimation of trees crown volume by 3D laser scanning system," in *2011 International Conference on Computer Distributed Control and Intelligent Environmental Monitoring (CDCIEM)*, February 2011, pp. 265–268.

[19] X. Zhang, J. Morris, and R. Klett, "Volume measurement using a laser scanner," Communication, and Information Technology Research (CITR) Computer Science Department, The University of Auckland, Tech. Rep., 2005.

[20] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, November 2000.

[21] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, ser. CVPR. Washington, DC, USA: IEEE Computer Society, 1997, pp. 1106–1112.

[22] P. D. Kovesi, "MATLAB and Octave functions for computer vision and image processing," Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia, available from: <http://www.csse.uwa.edu.au/∼pk/research/matlabfns/>.

[23] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.

[24] Y. S. Heo, K. M. Lee, and S. U. Lee, "Robust stereo matching using adaptive normalized cross-correlation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 807–822, April 2011.

[25] A. Donate, X. Liu, and E. Collins, "Efficient path-based stereo matching with subpixel accuracy," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 1, pp. 183–195, February 2011.

[26] W. Li, H. Yao, R. Ji, P. Xu, X. Liu, and D. Zhao, "Robust stereo matching combining SIFT descriptor with NCC under MRF framework," in *2010 First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA)*, September 2010, pp. 1018–1021.

[27] Z. Wang and Y. Quan, "An improved method for feature point matching in 3d reconstruction," in *International Symposium on Information Science and Engineering (ISISE)*, vol. 1, December 2008, pp. 159–162.

[28] R. Labati, A. Genovese, V. Piuri, and F. Scotti, "Fast 3-D fingertip reconstruction using a single two-view structured light acquisition," in *2011 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS)*, September 2011, pp. 1–8.

[29] N. Moshtagh, "Minimum volume enclosing ellipsoids," *GRASP Laboratory, University of Pennsylvania*, 2009, unpublished note.

[30] R. Guerchouche and F. Coldefy, "Camera calibration methods evaluation procedure for images rectification and 3D reconstruction," in *Proceedings of the 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG'2008)*, February 2008, pp. 205–210.

[31] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*, 2nd ed. Wiley-Interscience, November 2001.