Volume Measurement of Human Hand by Using Digital Image Sequences

Mahdi Arghiani, Ehsan Nemati, Hadi Sadoghi Yazdi

Engineering Department, Tarbiat Moallem University of Sabzevar, Sabzevar, Iran

Arghiani@eng.salamgroup.net

Nemati@eng.salamgroup.net

sadoghi@sttu.ac.ir

Abstract: In this paper, we introduce the new method for volume measurement of rigid objects. A machine vision algorithm is developed which estimates human hand volume from two-dimensional digital images that captured from different views. The proposed algorithm is general and can easily be used for other objects. The novelty of our method lies on the use of ordinary devices, simple algorithm for implementation, and high speed in running program. Main idea includes volume measuring using projection of object image from different views. Error compensation in object detection and feature extraction is performed using suitable estimators such as adaptive neuro fuzzy inference system and support vector regression. Ability of the proposed system is studied in volume measurement of box and human hand.

Keywords: Image Processing; Volume Measurement; Human Hand; Adaptive Neuro Fuzzy Inference System; Support Vector Regression.

1. Introduction

A volume for non-geometric objects plays the important role in many applications. For example, the left ventricular volume has the utmost importance to determine the efficiency and applicability of diagnostic technique [1]. Other applications of it, are measuring the dimensions of skin wounds [2], and a volume of agricultural product that able us to calculate the mass and density [3, 4]. Thus, it is a valuable work to constructing a system for volume measuring. To satisfy all applications, it requires an appropriate method. The variety of volume calculation systems are proposed, but each one is suitable in the limited range of application. Some of those are including the liquid displacement method [3, 4], gas displacement method, stereoscopy and stereo vision [19], resonance frequency [18], 3-D ultrasound image [6, 9], photographic tomography using structured light [5, 8] and using range sensors [11].

The volume measurement system was explored for a long time. The liquid displacement method is a simple and easy way, but for instance, agricultural products or foodstuffs may be damaged by their immersion into liquid [3]. The gas displacement method too, has its advantages and drawbacks. One of its difficulties is that takes a long time for measurement. Stanley and et. al. [5] measured the surface and area of the human body with structured light that need extra equipment and lead to high computational costs. Schmitt determined the limb volume by using of ultrasound [6]. Determining the volume of objects with this method, require many devices. In this case, they used transducers, amplifiers, transmitter, receiver and signal generator. Cheong and et. al. [1] measured the Tibial Cartilage volume by means of magnetic resonance imaging (MRI) that has its disadvantages. Another method is simulated from human visual system called Stereoscopy [7] that using the image with two different angles for construction 3-D objects.

As mentioned in the above papers some problems exist such as, using extra devices (high technology) and dealing with high computational cost in the process of volume measurement. In this paper, we introduced a novel approach to measuring the volume of human hand based on capturing image of hand from different views. This goal obtains by one cheaper digital camera, simple algorithm for implementation, and high speed in running program. For testing the capability and reliability of our method, we apply our system to measure the volumes of boxes, and then we extend the system to measuring the human hand volume. Results for both experiments describe and discuss in the next sections with following organization.

Section 2 appropriates to explain the proposed system including capturing image, preprocessing, feature extraction, and regressor scheme. Experimental results are discussed in section 3. This

section pays to describing the method of data collection. Also, results of each part of the proposed system are checked and obtained errors are studied. Finally, conclusion and future works are presented in section 4.
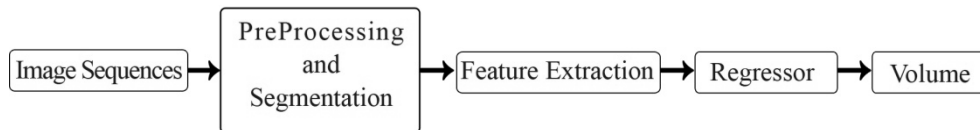
2. The proposed method

Our method is similar to the stereoscopy and stereo vision with some differences. We capture objects from variety of views only with one digital camera, in contrast to using camera arrays [19]. Furthermore, we don't involve in obtaining 3-Dimensional models because of their complicated algorithm and computations. The volume measurement flow chart is shown in the Fig. 1.

Fig. 1: The proposed system

Image sequences are images that captured from different views of same object. These views (orientations) must be the same for all different objects. Wide changes in these views affect the system performance. Besides, these views must reflect all features of object such as eminence and notch. For example, suppose there are two similar boxes that in one of them exists a hole. So capturing operation must contain the effect of hole in the measurement of box volume. Otherwise, the system shows the same volume for two boxes. Of course, we try to lessen the number of captured images to boost the system speed. Nonetheless, it is a trade off between precise and speed.

In the pre-processing and segmentation section, thresholding, and region labeling are accomplished. Segmentation is one of the important tools in image processing. Various algorithms have been proposed that we can find them in image processing books [13-17]. In this paper, we use segmentation algorithm based on fuzzy c-means algorithm which is a clustering technique. This algorithm is explained in the Appendix part a. After segmentation, object is extracted using labeling and number of pixels is applied to regressor as feature. The extracted features from images depend on the shape of objects. Then, a regressor calculates the volume of objects. This regressor approximates the function *f* that has the following form:

$$V = f\left(x_1, ..., x_n\right) \qquad (1)$$

Where $x_1, ..., x_n$ are number of pixels from different views of object. V is the volume of object in cubic centimeters. The *f* is a function that approximates the object volume from the input data. As a function, *f* is highly non-linear, but it may be approximated with some regression methods such as ANN[1], SVR[2], and ANFIS[3]. Despite some methods such as that used in [4], our function need not any proportional constant and straightly calculates the volume. Due to this subject, our method is speedier in running. To achieve the best results, we approximate this function with two methods: support vector regression and adaptive neuro fuzzy inference system.

2.1. Support vector regression

The support vector regression (SVR) is a supervised learning method that generates input-output mapping functions from a set of labeled training data. The mapping function can be either a classification function, i.e., the category of the input data, or a regression function. Initially developed for solving classification problems, support vector techniques can be successfully applied to regression. This learning method is explained in more details in the appendix part b. We try developing SVR for increasing robustness against the noise according previous work over support vector machine [20].

2.2. Adaptive neuro Fuzzy Inference System

Recently, there has been a growing interest in combining neural network and fuzzy inference system; as a result, neuro-fuzzy computing techniques have been evolved. Neuro-fuzzy systems are fuzzy systems, which use neural networks theory in order to determine their properties (fuzzy sets and fuzzy rules) by processing data samples. Neuro-fuzzy integrates to synthesize the merits of both neural networks and fuzzy systems in a complementary way to overcome their disadvantages. Some details over this method are appeared in the appendix part c.

In the next subsection, we explain the process of measuring the volume for two objects: box and human hand.

2.3. Box Volume Measurement

---

[1] Artificial Neural Network
[2] Support Vector Regression

[3] Adaptive Neuro Fuzzy Inference System

We know that volume of box obtain from formula $V = a \times b \times c$, where a, b and c are the box sides. If all boxes (of different volumes) are photographed in exactly same orientations, then the box volumes would be a monotonic function of the areas extracted from images.

The black color of all boxes in white background helps the segmentation algorithm to work better. To eliminate the object shadows, we use three lights from three different points. Capturing conditions for all boxes must be the same. These conditions are such as distance from camera, capturing orientation and number of images for each box. We capture them from four different views:

    i.   With respect to greater face
    ii.   With respect to middle face
    iii.   With respect to smaller face
    iv.   With respect to one fixed corner

These four images are shown in fig. 2:
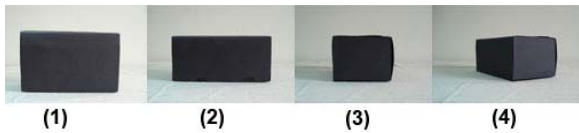

Fig. 2: A box from different views

Then pre-processing and segmentation are applied to the fig .2. Resultant image showed in fig. 3.
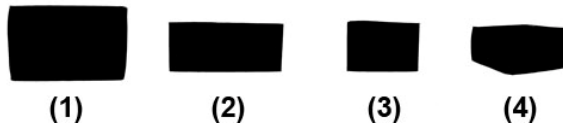

Fig. 3: Box images after segmentation

Our scrutiny on the errors with and without image number 4 in fig. 2 shows that this view is not very effective and the results are the same as state which this view doesn't exist.

The features extracted from these images are areas measured for every view. These features are same as x_1 to x_4 in equation (1). Proportion of real volume and measured volume using the proposed approach must be equal approximately, for all boxes. However, due to errors in making boxes, this ratio varies for each box. Fig 4 depicts these variations.
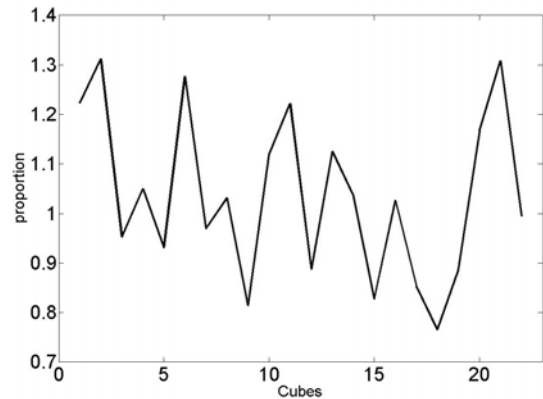

Fig. 4: Proportion of real volume and measured volume using the proposed system

2.4 . Human Hand Volume Measurement

After boxes, we apply the proposed system to measure human hand volume. For obtaining learning data, the volume of human hand is measured using variation of water volume (water displacement method) as shown in Fig 5. To lessen the errors in this stage, we measure ten times every hand volume and calculate its average to learn regressor.

Fig. 5: Measuring hand volume for regressor training

Due to variety in color of different hands, every person wears white glove. To capture the images of hand, we define seven points of view as shown in Fig. 6.
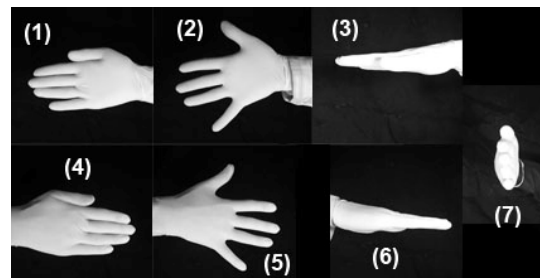

Fig. 6: Hand from different views

The grayscale images are obtained after applying pre-processing and segmentation procedure. Fig 7 shows results in the example of Fig 6.
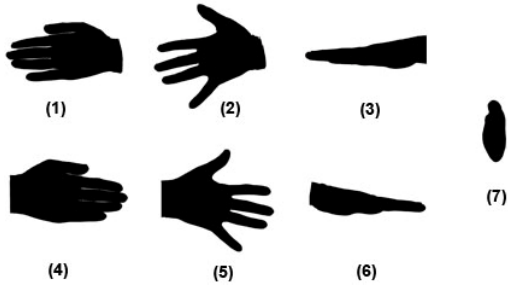
Fig. 7: Hand images after pre-processing

In this case, the inputs of function in equation (1) are x_1 to x_7 which are the areas of hand in each image. The images 2 and 5 are similar to images 1 and 4 but with different boundary. So it is better that they do not eliminate.

3. Experimental results

Two experiments over volume measuring of boxes and human hand are performed for idea checking. 22 boxes are constructed and four different views are selected for capturing images. Also for measuring of hand volume, 11 human with different hand volumes are selected. 77 images from 7 views are collected.

Different errors occur in these experiments such as,

✓ Errors in constructing boxes
✓ Errors in segmentation of human hand and boxes
✓ Errors in water displacement method
✓ Little number of training samples.

These errors influence the obtained results. SVR and ANFIS are trained over $N-1$ samples and are tested over one remaining sample. Test sample change N times. Obtained errors is different between main volumes (it is measured for hand according to Fig 5 and for boxes with a geometrical relation) and obtained volume from trained SVR and ANFIS systems over N test samples. Fig 8 and 9 shows error for measuring of volume for boxes using SVR and ANFIS respectively.
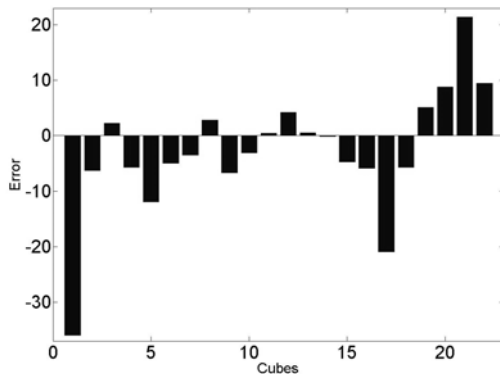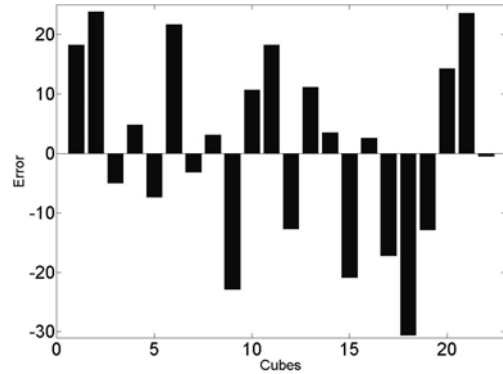

Fig. 8: Error of SVR for boxes

Fig. 9: Error of ANFIS for boxes

The best applied kernel for SVR is obtained by RBF kernel with variance σ^2 equals to 4 according as follows,

$$K(x, y) = \exp\left(-\frac{\|x - y\|}{2\sigma^2}\right) \qquad (2)$$

σ^2 must increase for small number of training samples however error increases in testing procedure. Difference in structure of SVR and ANFIS shows different errors over testing samples. Therefore, fusion of results can help to obtain lower error. With equal effect, volume is measured as follows and obtained error showed in Fig 10.

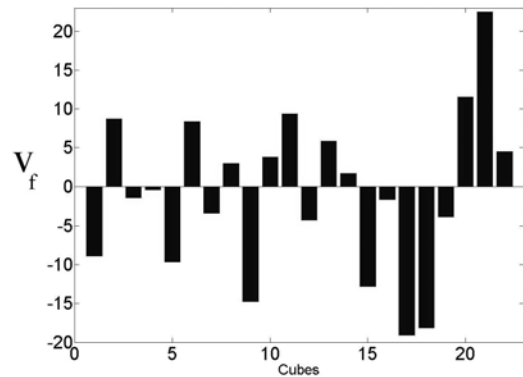$$V_f = \frac{1}{2}V_{Anfis} + \frac{1}{2}V_{SVR} \qquad (3)$$

Fig.10: Result (Error) of ANFIS and SVR fusion

Similar work performs over measuring of hand volume. Fig 11 and 12 shows error over test samples. Fig 13 shows results of mixed system.
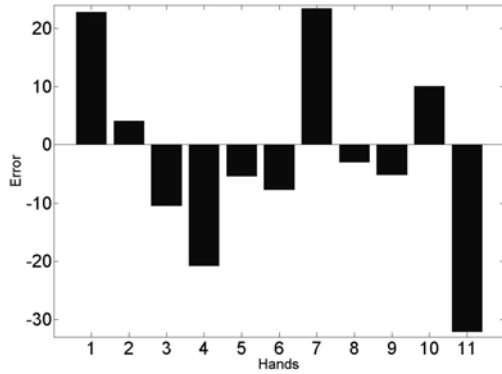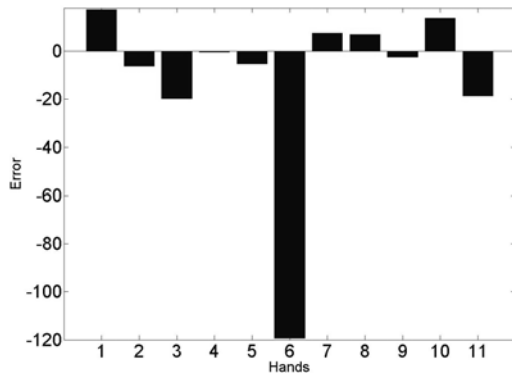
Fig. 11: Error of SVR for human hand

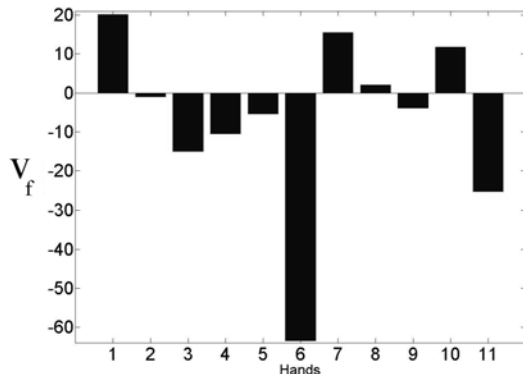**Fig. 12: Error of ANFIS for human hand**

Fig. 13: Result (Error) of ANFIS and SVR fusion

4. Conclusion and future work

An algorithm, which estimates the volume of geometric and non-geometric objects, has been developed. The algorithm is based on the number of pixels of object from different views. Because of the proposed scheme is not based on the geometrical features (length, width, height and so on), it should be able to applied to different objects. The regressor can be retrained to obtain volume of almost every object. 8.06% average error over measurement of 22 boxes and 15.83% average error over measurement of 11 human hands showed applicability of the proposed scheme.

For achieving better results with lower error, we will contrive to improve the environment illumination and camera parameters. We intend to extend our database for hand volume measurement. With greater database, the regressor will lead to better results. In addition, with finding better features of objects, as boundary and so on, we will enhance regression. As mentioned the capturing orientation is very important and we will search for better and more reliable capturing orientations. Also, we intend to apply our method to measuring the human body volume.

Appendix

Part a) *Fuzzy c-means algorithm*

One of the most widely used fuzzy clustering models is fuzzy c-means (FCM). The FCM algorithm assigns memberships to which are inversely related to the relative distance of to the point prototypes that are cluster centers in the FCM model. Some problems in FCM are as follows,

a) Samples with equidistance to centers
b) Measurement of distance to crisp centers
c) Data are crisp

Objective function in FCM is

$$J_m(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^m \left\| x_k - v_i \right\|^p \quad \text{(a-1)}$$

Where $V = \{v_i, i=1,2,...,c\}$, $U = \{u_{ik}^m\}$ are centers and membership functions and $X = \{x_1,...x_n\} \subset R^S$ are data. $v_i \in R^s$ is center of i_{th} centers. $u_{ik} \in [0,1]$ is membership of i_{th} data to k_{th} centers. N samples are clustered to c cluster as following constraints are satisfied.

$$M_{fcm} = \left\{ U \in R^{c*n} \middle| \begin{array}{l} \forall i,k: \circ \leq u_{i,k} \leq 1 \ ; \\ \sum_{i=1}^{c} u_{ik} = 1 \ , \ \sum_{k=1}^{n} u_{ik} > 0 \end{array} \right\} \quad \text{(a-2)}$$

Optimization procedure gives,

$$v_i = \frac{\sum_{k=1}^{n} (u_{i,k})^m x_k}{\sum_{k=1}^{n} (u_{i,k})^m} \quad \text{(a-3)} ,$$

$$u_{ik} = \sum_{l=1}^{c} \left(\frac{\left\| Y_k - v_i \right\|}{\left\| Y_k - v_l \right\|} \right)^{-2/(m-1)} \quad \text{(a-4)}.$$

Part b) *Support vector regression*

Suppose we are given training data $\{(X_1,y_1),(X_2,y_2),...(X_l,y_l)\} \subset X \times R$, where X denotes the space of the input patterns (e.g. $X = R^D$). In ε-SV regression [Vapnik, 1995], our goal is to find a function f(x) that has at most ε deviation from the

actually obtained targets y_i for all the training data. The regressor must not only fit the given data well, but also make minimal errors in predicting the values at any other arbitrary point in R^D. Nonlinear regression is accomplished by fitting a linear regressor in a higher dimensional feature space. A nonlinear transformation ϕ is used to transform data points from the input space (with dimension D) into a feature space having a higher dimension L ($L > D$). The nonlinear mapping is denoted by $\phi : R^D \rightarrow R^L$.

This problem can be written as a convex optimization problem; hence, we arrive at the formulation stated in [Vapnik, 1995].

$$Min \quad \frac{1}{2}\|W\|^2 + C\left(\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right)$$

$$s.t \quad y_i - W^T\phi(X_i) - b \le \varepsilon + \xi_i \qquad \text{(b-1)}$$
$$- y_i + W^T\phi(X_i) + b \le \varepsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0$$

Where C > 0 is a constant, ξ_i, ξ_i^* are slack variables for soft margin SVM, which allow accepting some deviation larger than ε that is precision. It turns out that in most cases the optimization problem (b-1) can be solved more easily in its dual formulation.

$$Max \quad -\frac{1}{2}\sum_{i,j=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)K\left(X_i, X_j\right) -$$

$$\varepsilon\sum_{i=1}^{l}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{l}y_i\left(\alpha_i - \alpha_i^*\right) \qquad \text{(b-2)}$$

$$s.t \quad \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0 \quad , \quad \alpha_i, \alpha_i^* \in [0, C]$$

Where the α_i, α_i^* are Lagrange coefficients and matrix K is termed as a kernel matrix and its elements are given by

$$K\left(X_i, X_j\right) = \phi(X_i)^T\phi(X_j) , \quad i, j = 1, 2, ... M .$$

By solving (b-2) we can find Lagrange coefficients and by replacing them, we have

$$W = \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) \times \phi(X_i) , \text{ thus}$$

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K\left(X_i, X_j\right) + b \qquad \text{(b-3)}$$

Part C) *Adaptive neuro Fuzzy Inference System*

ANFIS has been proved to have significant results in modeling nonlinear functions. In an ANFIS, the membership functions (MFs) are extracted from a data set that describes the system behavior. The ANFIS learns features in the data set and adjusts the system parameters according to given error criterion. In a fused architecture, NN learning algorithms are used to determine the parameters of fuzzy inference system. Below, we have summarized the advantages of the ANFIS technique.

- Real-time processing of instantaneous system input and output data. This property helps using this technique for many operational researches problems.
- Offline adaptation instead of online system-error minimization, thus easier to manage and no iterative algorithms are involved.
- System performance is not limited by the order of the function since it is not represented in polynomial format.
- Fast learning time.
- System performance tuning is flexible as the number of membership functions and training epochs can be altered easily.
- The simple if–then rules declaration and the ANFIS structure are easy to understand and implement.

References:

[1] J Cheong, D Suter, F Cicuttini "Development of Semi-automatic Segmentation Methods for Measuring Tibial Cartilage Volume." Digital Image Computing: Techniques and Applications, Dec. 2005.

[2] B. F. Jones, P. Plassmann "An Instrument to Measure the Dimensions of Skin Wounds." IEEE Trans. on Biomedical Engineering. vol. 42. no 5. pp. 464-470. may 1995.

[3] T. Charoenpong, P. Kamhom, K. Chamnongthai, M. Krairiksh "Volume measurement of mango by using 2D ellipse model". IEEE International Conference on Industrial Technology (IClT). Dec. 2004. pp. 1438-1441.

[4] K A Forbes, G M Tattersfield "Estimating fruit volume from digital images". Sept. 1999 pp.107 – 112. vol. 1.

[5] S. M. DUNN, I. L. KEIZER, J. YU "Measuring the area and volume of the human body with structured light". IEEE Trans. on systems, man, cybernetics, vol. 19, no. 6, November/december 1989. pp. 1350-1364.

[6] Schmitt, R.M.; Baisch, F.W.; Klein, M.; Welsch, H.J.; Becker, P.; "Limb volume determination by means of Ultrasound." Engineering in Medicine and Biology Society, Proceedings of the Annual International Conference of the IEEE. Nov. 1988. pp.481 – 482. vol.1

[7] Dhond U. R., aggarwal J. "Structure from stereo – A review." IEEE Trans. on syst. Man and Cybern. Vol. 19, 1989. pp. 1489-1510.

[8] Q Yang. "Finding stalk and calyx of apples using structured lighting. Computers and Electronics in Agriculture", 1993.

[9] A Laing, Q Smit, B Mortimer, and J Tapson. "Ultrasonic fruit sizing device". Journal of the South African Acoustics Institute, 1995.

[10] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms". New York: Plenum, 1981.

[11] K. Yarnane, S. Tanaka. "Volume Measurement of the Pulverulent Body Carried by a Conveyor Belt Using Range Sensors". SICE. Aug. 5-7. 2002. pp. 2936-2938.

[12] R. Ishiyama, T. Okatani, K. Deguchi. "Precise 3-D measurement using uncalibrated pattern projection." ICIP 2007 pp. 225-228.

[13] M. S. Nixon and Alberto S. Aguado "Feature Extraction and Image Processing.", Newnes 1st edition, pp 291. ISBN: 0750650788.

[14] David A., F. Jean, "Computer Vision - A Modern Approach." Prentice Hall, pp 401. ISBN: 0130851981.

[15] "Handbook of Mathematical Models in Computer Vision." Springer, 1st edition, ISBN: 0387263713

[16] William K. Pratt. "Digital Image Processing: PIKS Inside". 3rd Edition. John Wiley & Sons, Inc. 2001.ISBN: 0-471-37407-5. pp 551.

[17] R. Lukac. K.N. Plataniotis. "Color Image Processing, Applications and Methods." CRC Press. ISBN: 0-8493-9774.

[18] T. Nishizu, Y. Ikeda, Y. Torikata, S. Manmoto, T. Umehara, and T. Mizukami, "Automatic, Continuous Food Volume Measurement with a Helmholtz Resonator." Agricultural Engineering International: the CIGR Journal of Scientific Research and Development. 2004. Vol. III.

[19] P. Witoonchart, Richards Foulds, "Three-Dimensional Surface and volume Measurement Using a Camera Array." IEEE 28th Annual Northeast Bioengineering Conference, pp. 145 – 146, April 2002.

[20] Hadi Sadoghi Yazdi, S. Effati, Z. Saberi, "The Probabilistic Constraints in the Support Vector Machine," Applied Mathematics and Computation, (2007), doi:10.1016/j.amc.2007.04.109.