# Contribution in the Analysis and Coding of Three-Dimensional Image Sets

Nicholas P. Sgouros

Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
Panepistimiopolis, Ilissia
Athens 15784, Greece

nsg@di.uoa.gr

**Abstract.** Three-Dimensional (3D) visualization systems are used in many specialized applications where 3D observation is required. Technological achievements in the areas of integrated optics, sensors and network infrastructures guarantee that these systems will be used soon in a large number of different applications. This dissertation summarizes the author's research on the major aspects of a specific 3D visualization method, called Integral Photography (IP). This work describes the implementation of a prototype device capable of producing high resolution, near field, IP images of real 3D objects based on a flatbed scanner. In addition, this work also presents a novel automated method for calibration of the sensors and the optics used in an IP acquisition device without prior knowledge of the system's characteristics. This is accomplished by post-processing of the acquired IP images using image analysis and pattern recognition methods. Finally two encoding schemes are developed in order to properly exploit the inherent redundancy of an IP image and achieve high performance in a rate distortion sense.

**Keywords:** Three-dimensional Imaging, Integral Photography, Three-dimensional Image Analysis, Three-dimensional Image coding, Autostereoscopic Applications

## 1 Introduction

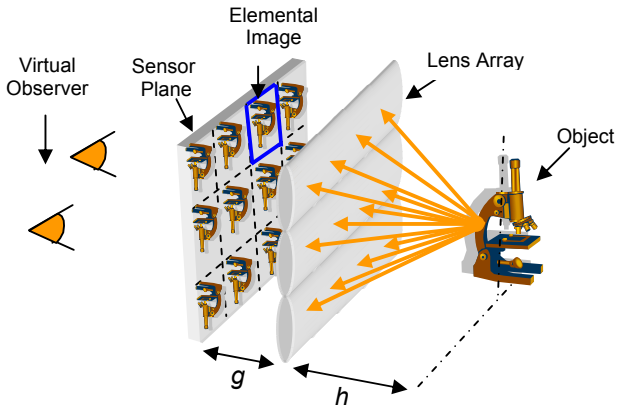
In general, three-dimensional (3D) viewing methods can be divided in stereoscopic and autostereoscopic. Stereoscopic methods require additional viewing aids to provide 3D viewing, like glasses or helmets, while autostereoscopic methods do not require additional viewing aids as all optical components needed in order to reproduce correct 3D objects and scenes are integrated in the display device [1-2].

Integral Photography (IP) is an autostereoscopic technique for producing high realistic 3D images, initially devised by G.Lippman [3] back in 1908. Specifically, this technique provides natural viewing with full colour support, enhanced detail and adequate depth level. Moreover, multiple simultaneous viewers are supported, while most of the currently existent IP setups for capturing and display provide full parallax

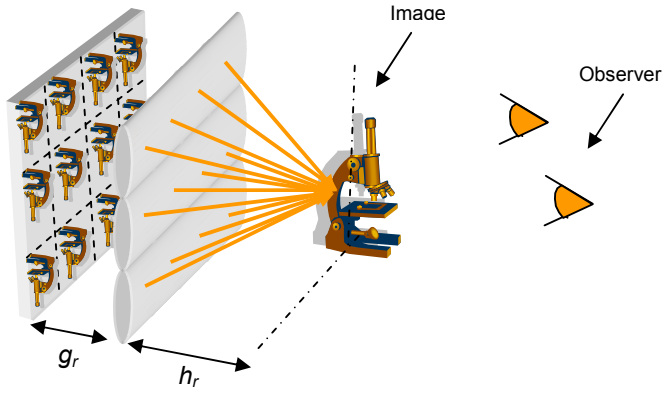
---

Dissertation Advisor: Manolis Sangriotis, Assoc Professor

on both horizontal and vertical directions [4,5]. The basic principle for the capturing and reproduction methodology is depicted in Fig.1, (a) and (b) respectively.




(a)




(b)

**Fig. 1.** IP (a) Acquisition and (b) Display Device functionality principles.

An IP image is a two-dimensional (2D) lattice of small images, usually called elemental images, which are formed behind each lens of the lens array of Fig. 1a. The IP image is captured using a Charge Coupled Detector (CCD) placed directly behind the lens array located at the sensor plane in the same figure. When the resultant IP image is displayed using a Liquid Crystal Display (LCD) through a second lens array, as shown in Fig.1b, the initial 3D scene is reproduced.

Some of the key problems of an IP imaging system involve calibration, image restoration, compression, depth extraction, object recognition and others. Most of the standard image processing techniques can be applied in resolving these issues. However, certain adaptations should be made in order to take advantage of the inherent characteristics of an IP image structure and produce optimal results. Under this framework there are a number of attempts made, trying to tackle problems such as IP image compression [6,7], but there are still major processing issues, like the ones mentioned, that remain an open field for research and development.

This paper is organized in four sections. In section one, a novel architecture for high quality IP image acquisition is proposed which uses a Lens Array and a commercial flatbed scanner. General calibration of the acquisition device along with filtering for noise removal and IP image analysis issues are resolved in section two. Finally in section three of this work two novel coding schemes are presented for IP and generalized multiview architectures. Section four, concludes this paper and presents future work issues and enhancements.

## 2 IP Still image Acquisition Device

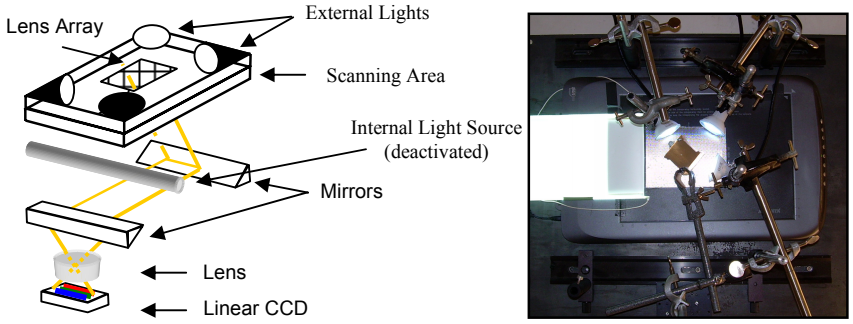
There are currently numerous IP acquisition setups based on digital sensors like area CCDs' [8-10]. The small size of the CCD sensor used in standard digital cameras imposes a number of limitations to IP acquisition setups. As a result, the IP images produced have a small number of lenses in an effort to increase lateral image resolution. On the contrary, an increase to the number of lenses in order to enhance depth resolution reduces lateral resolution. Taking into account both previous limitations a compromise is usually made based on the camera characteristics and available lens arrays. In this paper an IP acquisition device is described which is based on the functionality of a flatbed scanner and a large area lens Array. This device delivers high-resolution IP images with only a small fraction of the cost of solutions that are based on area CCD sensors.

### 2.1 Acquisition device

The developed prototype consists of a flatbed scanner with an effective scanning area of 21x30 cm and 0.1cm pitch for the rectangular lenses that assemble the lens array. The schematic of the device along with the functional prototype developed in this work are depicted in Fig. 2 (a) and (b) respectively.

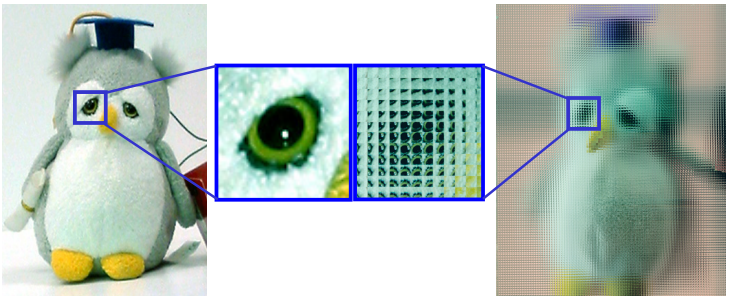
The optical resolution of the flatbed scanner used in the proposed setup is 3200dpi x 3200 dpi. When used in conjunction with a lens array of 0.1cm lens pitch high-quality IP images were produced, where each elemental image contained over 100 pixels in each dimension. The internal light source of the scanner was deactivated and external lighting was used in an effort to eliminate back-scattering occurring due to the incident light from the internal light source on the lens array. The only drawback of this device is the time needed for scanning the IP image of a 3D object, which limits its use to static objects. However, the resultant images unveil the potential of IP as a forthcoming standard for 3D acquisition and display. A high quality IP image,

along with the real world 3D object are depicted in Fig. 3 (a),(d). A magnification of a small part of the IP image in Fig. 3 (c) shows the rectangular structure of the lens array and the different views of the part of the object in Fig. 3 (b) in neighboring elemental images.




(a) (b)

**Fig. 2.** Acquisition Device: (a) Operation Principle (b) Experimental prototype.



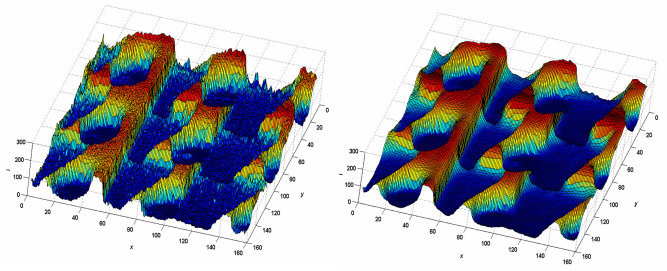
(a) (b) (c) (d)

**Fig. 3.** (a) Real world 3D object. (b) Magnified part of a portion of the object. (c) Corresponding IP image of the magnified part of the object in (b). IP image of the object in (a).

## 3 IP image Analysis

In this section, we develop a post acquisition filtering stage for noise reduction and propose a scene invariant method that automatically detects small rotational misalignments in IP image acquisition setups. In addition the technique described manages to measure the exact size and position of each elemental image without requiring prior knowledge on the system characteristics. In this way we can perform post acquisition processing in IP images or assist in the initial calibration of an IP image acquisition setup.

A Contrast-Limited Adaptive Histogram Equalization – CLAHE method is used in order to compensate for bad lighting conditions and a non linear 2D median filtering is applied in order to reduce excessive shot noise form the IP image. The results of the de-noising operation are depicted in Fig. 4.



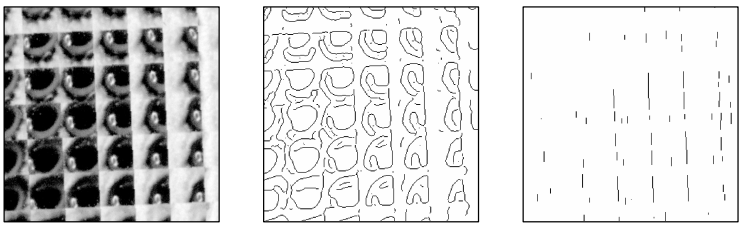
(a) (b)

**Fig. 4.** (a) Intensity values for a part of an IP image, (b) Intensity values for the same image after non linear 2D median filtering.

The rest of this section focuses on the accurate measurement of the skew angle of the Lattice Lines Structure (LLS) that reflects the rotational misalignment of the lens array used in the acquisition stage, in regard to the CCD sensor. It also accurately detects the lattice lines positions, which usually deviate from the equidistant case due to the non-integer values of the ratio of the lens array pitch to the CCD pitch. For this purpose, a non-linear filtering technique is used to enhance the boundaries of the elemental images along with the Hough Transform (HT) [11] in order to produce an accurate estimate of the skew angle. A lattice matching method is developed, based on the techniques used to solve the longest common subsequence (LCS) problem [12], in order to match a detected sequence of lines positions to a theoretical lattice model. The experimental data consist of a series of uncalibrated IP images acquired using the setup proposed in section two and computer generated IP images, using the method proposed in [13].

### 3.1 Skew Detection

A color IP image is initially converted to gray-scale as shown in Fig. 5a and the Canny edge detector [14] is deployed next in order to detect strong lines in the IP image and produce the edge image depicted in Fig. 5b. The image-wide lattice lines in the edge image are further enhanced using a one-dimensional (1D) median filter. On the contrary, shot noise and lines running to other directions are effectively attenuated due to their small size in regard to the median filter window. The filter is applied row-wise or column-wise and has the effect of attenuating everything except line segments that are almost horizontal or vertical respectively. In order to reduce the overall computational complexity of the algorithm, only one set of lattice lines, either horizontal or vertical, is considered in the skew detection process. As there is always a large number of elemental images in an IP image, there is also an adequate number of correctly detected lines running in a specific direction that is enough to produce an accurate estimation of the skew angle. The results of a vertical 1D median filtering are depicted in Fig. 5c.



(a) (b) (c)

**Fig. 5.** A portion of a physically acquired IP image. (a) Initial gray-scale image, (b) edge detection result, (c) edge image after vertical 1D median filtering.

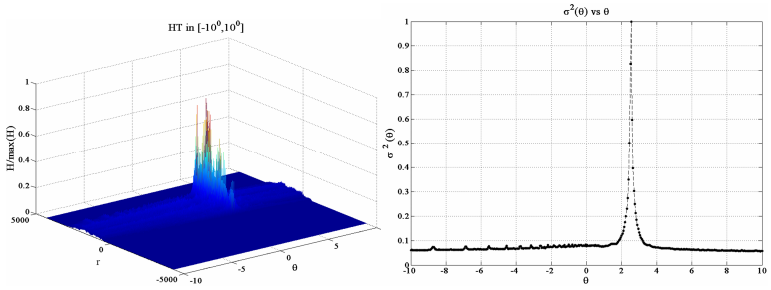
In the skew angle detection process the HT is used, which is a robust and extensively used technique in image processing applications. The HT succeeds in identifying only image-wide lines, which are more likely to correspond to true lines of the LLS that define the boundaries of the elemental images. In our framework we used a sampling rate of $0.05^0$ which provides a high level of accuracy, and assumed that the skew angle of the IP image remains under $\pm 10^0$ in order to reduce the computational complexity of the algorithm.

The detection of the skew angle ($\theta_s$) in the HT space is based on the fact that $\theta_s$ corresponds to a column in the HT accumulator array, **H**, having a large number of strong peaks, as shown in Fig. 6a. On the contrary the rest of the columns in **H** are relatively homogenous having a small number of weak peaks. In order to determine $\theta_s$ the statistical variance of each column in **H** is calculated and the results are shown in Fig. 6b.

### 3.2 LLS detection stage

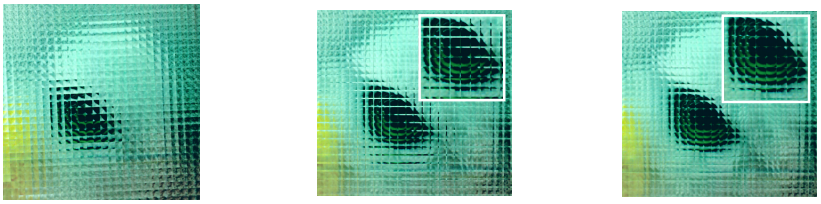
In order to derive the lattice lines positions that form the LLS the Canny edge detector and the 1D median filter are reapplied on the deskewed version of the initial gray-scale IP image. Since the image is deskewed, the lattice lines are properly oriented in horizontal and vertical directions. In order to enhance these image-wide lines the 1D median filter is applied row-wise and column-wise and the horizontal and vertical projection profiles [15] are calculated.

Peaks that are candidates for lattice lines in these two directions are identified using a global threshold at 20% of the maximum value of the corresponding profile. For each of the candidate peaks, a 1D version of the peak detection algorithm presented in [15] is used. The mean size of each elemental image is derived by calculating the average distance of all peaks detected in the previous stage in each direction. Finally a matching procedure utilizing the Smith-Waterman algorithm [16] follows that matches the detected LLS to a modeled LLS based on the mean size of each elemental image. The results of this procedure are shown in Fig. 7 as an application to pseudoscopic elimination [17].




(a) (b)

**Fig. 6.** (a) The column in the HT space corresponding to the skew angle $\theta_s$ has a large number of strong peaks, (b) the corresponding variance $\sigma^2(\theta)$ vs. $\theta$.



(a) (b) (c)

**Fig. 7.** (a) Uncalibrated portion of an IP image. Pseudoscopy elimination in the deskewed IP image, using (b) constant and (c) variable, pitch size.

## 4 IP and Multiview Image Coding

IP images are regarded as an omnidirectional multiview image set. In order to properly exploit both intra-pixel and inter-elemental image redundancy which is inherent in IP images two encoders are proposed. The first encoder uses disparity estimation between adjacent elemental images operating on the basic principles of the MPEG-2 compression standard [18]. The encoder outperforms previously proposed 2D coding schemes like JPEG for IP images containing medium and large sized elemental images. However, as disparity estimation schemes become impractical with small sized elemental images an adaptive 3D Discrete Cosine Transform (3D-DCT) is used that effectively exploits the intra-pixel and inter-elemental image redundancy by properly grouping neighboring elemental images.

### 4.1 Disparity encoder

The perfect alignment of subsequent lenses in the array allows a unidirectional search to be performed at the disparity estimation step, which has a significantly lower complexity compared to a multidirectional search to estimate a disparity vector. This feature also allows the use of a full pixel search, which further improves the matching results. The best matching for a predicted block is decided by the use of a mean square error (MSE) estimator.

Disparity vectors produced by the disparity estimation process are finally DPCM coded, followed by Huffman encoding that further reduces the extra bits needed for the disparity vectors. Finally, standard intra-frame and residual frame coding were used as described in MPEG-2 for coding reference and residual elemental images respectively. The results of the comparison with baseline JPEG are depicted in Fig. 8.

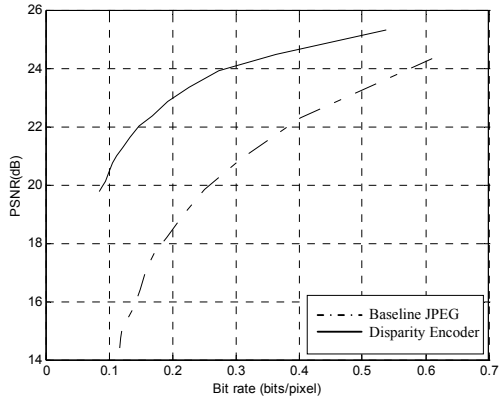



**Fig. 8.** Results for the Disparity Encoder vs Baseline JPEG.

### 4.2 Adaptive 3D-DCT encoder

The 2D lattice of elemental images is initially transformed to a 1D series of elemental images, using the a spiral curve as described in [7]. Consecutive elemental images are grouped together to form volumes on which the 3D-DCT will be applied. In our approach, the 3D-DCT is applied on groups of eight elemental images as a compromise between good quality and computational efficiency.

The encoder is assembled of a 3D-DCT unit, the quantizer and the entropy coder (EC). Two additional units are added to this standard setup used for detecting scene parts that belong to the background and determining the quantization values and the scan order of the 3D-DCT coefficients. The encoder layout is depicted in Fig 1.

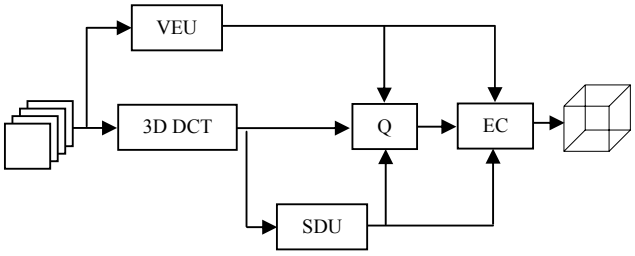



**Fig. 9.** Adaptive 3D-DCT Encoder.

The encoder uses a local variability strategy (VEU) on the initial data volume in order to point out parts of the image set that belong to the distant background. A global standard deviation strategy (SDU) is also used to locate the dominant 3D-DCT coefficients directivity and determine upon quantization values and scan order of the quantized coefficients. The performance results of the adaptive 3D-DCT scheme compared to baseline JPEG are depicted in Fig. 10.

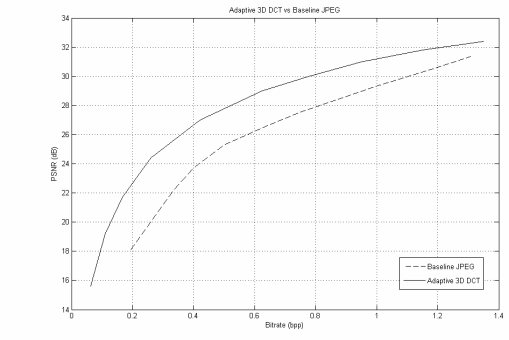



**Fig. 10.** Adaptive 3D-DCT vs Baseline JPEG.

## 4 Conclusions and Future Work

In this paper a holistic approach is presented for IP image acquisition, analysis, processing and encoding. The results showed that proper extensions of classic image analysis and processing algorithms achieve high accuracy and efficiency when applied on IP images. Future work includes 3D object reconstruction from IP images and development of algorithms and techniques for IP video acquisition and processing.

## References


1. S.Pastoor, M.Wöpking, 3-D displays: A review of current technologies, Displays 17 (2), pp 100-110, 1997.
2. M.Halle, Autostereoscopic displays and computer graphics, Computer Graphics 31(2), pp. 58-62, 1997.
3. G. Lippmann, La Photographie integrale, C.R. Acad. Sci. 146, pp. 446-455, 1908.
4. J. S. Jang, B. Javidi, Two-step integral imaging for orthoscopic three-dimensional imaging with improved viewing resolution, Opt. Eng. 41(10), pp. 2568-2571, 2002.
5. J. S. Jang, B. Javidi, Formation of orthoscopic three dimensional real images in direct pickup one-step integral imaging, Opt. Eng. 42(7), pp. 1869-1870, 2003.
6. N.Sgouros, A.Andreou, M.Sangriotis, P.Papageorgas, D.Maroulis, N.Theofanous, Compression of IP images for autostereoscopic 3D imaging applications, In: Proc. IEEE ISPA03, pp. 223-227, 2003.
7. S.Yeom, A.Stern, B.Javidi, Compression of 3D color integral images, Opt. Express 12(8), pp. 1632-1642, 2004.
8. M. Levoy, "Light Fields and Computational Imaging," IEEE Computer, vol. 39(8), pp. 46-55, 2006.
9. J. Jang and B. Javidi, "Time-Multiplexed Integral Imaging For 3D Sensing and Display," Optics & Photonics News, vol. 15, pp. 36-43, 2004.
10. M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca and G. Saavedra, "Integral Imaging with Improved Depth of Field by Use of Amplitude-Modulated Microlens Arrays," Applied Optics, vol. 43, pp. 5806-5813, 2004.
11. R.C.Gonzalez and R.E.Woods, *Digital image processing, second ed.*(Prentice Hall, NJ, 2002).
12. T.H.Cormen, C.E.Leiserson, R.L.Rivest, Introduction to algorithms, (MIT Press, MA, 2000).
13. S. S. Athineos, N. P. Sgouros et.al, "Photorealistic Integral Photography using a Ray Traced Model of the Capturing Optics," IS&T/SPIE Journal of Electronic Imaging, vol. 15(04), no. 043007, 2006.
14. J. F. Canny, "A Computational Approach for Edge Detection," Trans. Pat. Anal. Mach. Intell. 8, pp. 679-698, 1986.
15. R.C.Gonzalez, R.E.Woods and S.L Eddins, Digital image processing, using MATLAB (Prentice Hall, NJ, 2004).
16. T. Smith and M. Waterman, "Identification of common molecular subsequences," Journal of Molecular Biology, vol. 147, pp. 195-197, 1981.
17. M. Martinez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, "Formation of real, orthoscopic integral images by smart pixel mapping," Opt. Express 13, pp. 9175-9180 2005.
18. K. Rao, J. Hwang, Techniques & standards for image-video & audio coding, Prentice Hall,1996.