A SINGLE-CHIP REAL-TIME RANGE FINDER

A Dissertation

by

SICHENG CHEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2003

Major Subject: Electrical Engineering

A SINGLE-CHIP REAL-TIME RANGE FINDER

A Dissertation

by

SICHENG CHEN

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Uğur Çilingiroğlu (Co-Chair of Committee)	Edgar Sanchez-Sinencio (Co-Chair of Committee)
José Silva-Martínez (Member)	Duncan M. Walker (Member)
Chanan Singh (Head of Department)	

May 2003

Major Subject: Electrical Engineering

ABSTRACT

A Single-Chip Real-Time Range Finder. (May 2003)

Sicheng Chen, B.S., Huazhong University of Science and Technology;

M.S., Institute of Computing Technology, Chinese Academy of Sciences

Co–Chairs of Advisory Committee: Dr. Uğur Çilingiroğlu
Dr. Edgar Sanchez-Sinencio

Range finding are widely used in various industrial applications, such as machine vision, collision avoidance, and robotics. Presently most range finders either rely on active transmitters or sophisticated mechanical controllers and powerful processors to extract range information, which make the range finders costly, bulky, or slowly, and limit their applications. This dissertation is a detailed description of a real-time vision-based range sensing technique and its single-chip CMOS implementation. To the best of our knowledge, this system is the first single chip vision-based range finder that doesn't need any mechanical position adjustment, memory or digital processor. The entire signal processing on the chip is purely analog and occurs in parallel. The chip captures the image of an object and extracts the depth and range information from just a single picture. The on-chip, continuous-time, logarithmic photoreceptor circuits are used to couple spatial image signals into the range-extracting processing network. The photoreceptor pixels can adjust their operating regions, simultaneously achieving high sensitivity and wide dynamic range. The image sharpness processor and Winner-Take-All circuits are characterized and analyzed carefully for their temporal bandwidth and detection performance. The mathematical and optical models of the system are built and carefully verified. A prototype based on this technique

has been fabricated and tested. The experimental results prove that the range finder can achieve acceptable range sensing precision with low cost and excellent speed performance in short-to-medium range coverage. Therefore, it is particularly useful for collision avoidance.

To my parents and dear wife

ACKNOWLEDGMENTS

First of all I would like to express my sincere gratitude to my academic advisor Dr. Uğur Çilingiroğlu, for his intelligent guidance and constant encouragement during the research. I also wish to thank Dr. Edgar Sanchez-Sinencio for showing me what it means to be a scientist. I thank Dr. Uğur Çilingiroğlu, Dr. Edgar Sanchez-Sinencio, Dr. José Silva-Martínez, and Dr. Duncan M. Walker for serving as my thesis committee and for all the valuable advices. I also wish to thank Dr. John E. Bauer as the Graduate Council Representative.

I thank Dr. Aydin Karsilayan who manages the cadence design and PCB manufacturing environments. I also greatly appreciate Dr. Sherif HK Embabi and Dr. José Pineda de Gyvez for introducing me to the analog design during my first year in the group. I would like to thank my friends and colleagues from the Analog and Mixed Signal group, Sung Tae Moon, who gave me lots of help in board fabrication and chip testing; Siew Kuok Hoon, Tuli Dake, Anand Veeravalli, Asit Shankar, Bo Xia, who gave me great help during the first year. I thank Rodrigo Oliveira-Pinto, Ari Yakov Valero-Lopez, Shanfeng Cheng, and Taner Sumesaglam who maintain the NT and workstation network in the lab. I thank Chunyu Xin, Guang Song, Jozef Adut, and Xingyu Tang for their help on LaTeX document preparation. Many thanks also go to Michael McConnell, Sean McEnroe, and Jun Chen for their help during my internship in Texas Instruments Inc.

I especially thank Emre Çilingiroğlu and Dr. Donald Parker, for their great help and instructions on the mechanical works on the camera body. Without their help, the environmental testing setup can never be built. I also want to thank Steve Smith for help me solving many problems when I worked in his workshop.

My deepest appreciation and love go to my beloved wife Jin Wang. I thank her

for her sacrifice and love during these years.

Finally, I acknowledge the financial support of Texas Instruments Inc. and department of Electrical Engineering. Chip fabrication was provided by the foundry service MOSIS, without which none of this work would have been possible.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

Every minute, on average, at least one person dies in a crash [4]. Auto accidents also injured at least 10 million people in 2001, two or three million of them seriously. All told, the hospital bills, damaged property, and other costs will add up to 1-3 percent of the world's gross domestic product, according to the Paris-based organization for Economic Cooperation and Development. For the United States alone, the tally will amount to roughly US $200 billion. Although seat belts and airbags have saved millions of lives each year, the ultimate solution is to keep cars from smashing into each other.

One category of accidents is rear-end-collision accidents. In 1999 alone, 1.848 million rear-end-collision accidents were officially recorded in U.S. This represents almost one third of all crash type accidents, and 12 percent of all traffic-accident related fatality [5]. It is estimated that as 50% of these accidents could have been avoided by deploying on-board collision avoidance systems [6, 7, 8]. Since the major cause of these accidents is driver inattention or distraction, the collision-avoidance systems must be capable of detecting and warning the driver of potential hazards around the vehicle. The distance between the vehicles and the distance from vehicle to obstacles are very important for these collision avoidance systems.

Another important category of range sensing applications is robotics. To avoid obstacles in a three-dimensional workspace, robots need to have the real-time range information of the objects in it moving path.

[1]

[1]This dissertation follows the style and format of *IEEE Transactions on Automatic Control*.

Range finding also can be used in other industrial applications, such as machine vision. For example, a machine vision system with range finding capability can be used for machine safety. The mining equipment must have knowledge of the absolute or true position and orientation of the machinery in relation to the surroundings of the machinery.

A. Background

Range finding is so important to industry that until 2002 US Patent Office has approved more than 700 patents on different kinds of range finders.

Generally speaking, there are two ways to measure range: Contact and noncontact measurement. Contact measurement is very commonly used in everyday life. The distance to a point on an object is measured through a calibrated mechanical device that simultaneously connects the selected point to a reference position. For example, rulers and tapes can be used to measure the distance between two tables with the precision up to $1mm$, and a caliper is needed to measure the range between two screws with precision up to $0.1mm$, etc. The contact measurement is very cheap and it is both very accurate and precise. The chief disadvantage of mechanical approaches is that they are usually restricted to distances and work volumes up to a few meters at maximum. This is due to fundamental scaling laws for mechanical structures. Another disadvantage of contact approach is that they are not quick enough to measure distance in real-time. Therefore, contact approaches are unsuitable for most of the collision avoidance applications.

Only noncontact approaches are applicable in collision avoidance systems. Although it is difficult for the noncontact approaches to achieve very high precision, range finders used in collision avoidance systems usually do not need very high pre-

cision.

Precision and accuracy are the two most important specifications for range finders. Although they are often used interchangeably in common usage, in fact, they have quite different scientific meanings. The accuracy of a measurement refers to how well, within the limitations of the technique, the result corresponds to the actual value of the quantity being measured. The precision of a measurement refers to the extent to which the associated uncertainty has been minimized. For example: Suppose that the distance between two objects is $3cm$. If the range measurement is $3cm \pm 0.1cm$, it's accurate but not precise; if the range measurement is $3.5cm \pm 0.01cm$, it's precise but not accurate. Most range finder used in collision avoidance need to be very accurate but unnecessarily to be very precise. For example, in car collision avoidance system, it is acceptable to have the accuracy up to $10m \pm 1m$.

Noncontact distance measurement may be divided into active or passive techniques. Active techniques involve some forms of controlled energy (field or wave) linking a known reference location to the unknown target location. The source of energy is typically associated with the reference location, but in some cases the target, or both target and reference, may be active. Examples of this approach are sonar ranging based on time-of-flight of sound, and laser ranging and structured-light method based triangulation. Passive techniques rely on an externally occurring source of energy (e.g., sunlight or target/background temperature contrast) to make the target detectable.

In fact, automakers have already started equipping high-end vehicles with sensors that detect motion and obstacles, coupled to processors that respond instantly to whatever is detected [4]. These adaptive cruise control (ACC) systems, which add $1500 to $3000 to the cost of a car, use technologies including radar, IR laser and ultrasons to measure the distance of the vehicle ahead. The ACC system can be

found in Toyota's Progres compact luxury sedan, Nissan's Cima 41LV-2, Jaguar's XKR coupe, and some of Lexus's LS430 model. These are active sensors having both transmitting and receiving parts, and all operate on the principle of time of flight. Beside the cost, all of these active sensors suffer from interference problem.

Although laser range finders were considered in robotics navigation [9], they are not as widespread as one would hope. The main reason is that the requirements of workspace often lead to systems that are too costly and too bulky to be of practical use beyond research experiments. In addition, laser range finders are not suitable for robotics application in very short range ($< 2m$). Another major constraint for effective navigation is represented by time: When autonomous navigation is involved, the whole data processing has to be performed in real time. Moreover, in nonstructural environments, such as man-frequented areas, the employed range finder should be passive, avoiding the use of possibly harmful radiation (laser) and of artificial illumination of some kind, as well.

Vision is a natural choice for range sensing. Texture information at a fine angular resolution enables quite discriminative pattern recognition techniques. The human visual-perception system is perhaps the best example of how well such sensors might perform, if we add the appropriate processing. Besides, video cameras are cheap, and because they do not emit any signals, they raise no issues regarding interference with the environment [10].

Although vision-based technology is sensitive to weather conditions and dependent on natural or artificial illumination, it potentially offers the lowest cost because of the absence of any transmitter. Despite these advantages, though, no commercial application is yet available in this technology. This is at least partly attributable to the fact that classical vision-based techniques rely on multiple images of the same scene to extract range information, which makes them not only very costly in terms

of camera hardware, memory and signal processing requirements, but also very slow to be qualified as a real-time system. Another reason is that, vision-based range finding suffers from low illumination in bad weather. However, in robotic application, vision-based range finding is a good choice since the workspace luminance is usually high enough.

There are three categories of classical vision-based range-finding techniques [11]. One relies on stereovision, and extracts range information from the disparity of two slightly offset images of the same scene [5]. Although passive stereovision is one of the oldest research topics in the Computer Vision community, it is the most costly and the slowest method because it necessitates two separate cameras or complex view splitters and very heavy post processing. The major computational problems associated with stereo are the correspondence problem and detection of occlusion. Its use in robotics was limited by the large amount of computation required, the equivalent of dozens of correlation operation at each pixel. Beyond the computation issues, stereo suffers from limitations due to the triangulation geometry. One example of real-time stereo system in robotics is the SVM system from SRI [12], which uses computation on a conventional PC to deliver 320×240 range images at 12 Hz with 16 disparity levels.

Another category relies on "depth from defocus," (image defocus analysis or DFD) in which, multiple images of the same scene are registered under different lens parameters, and range information is extracted from the degree of blur in each. It measures the amount of blur in an image and needs as few as only two images to obtain the depth map of the scene. This idea is originally proposed by Pentland in [13, 14]. Pentland proposed a method that based on modeling a blurred step edge as the result of convolving a focused image with a Gaussian Point-Spread Function (PSF). He solved for the blur parameter and the height of the edge by a linear regression method to obtain the depth of the edge. He also proposed an algorithm for

an arbitrary scene using two images. One of the images is a focused image formed by a pinhole camera, and the other is a defocused image obtained from a wide aperture camera. This image is defocused by different amounts at different positions depending on the distance of object points. He then obtained the 3D information by comparing the corresponding points in these two images and measuring the change in focus.

All techniques of this category involve mechanical modulation of lens parameters, and require complex memory and signal processing hardware. Furthermore, they need an exact optical model of the imaging system.

The final category of vision-based techniques is known as "depth from focus," (image focus analysis or DFF) in which multiple images of the same scenes are acquired and are compared for the degree of focus. Range information is extracted from those lens settings that yield the best-focused image. Here, a large number of images are needed as inputs to compute a focus measure in order to determine the focused image. Reliance on a single camera and avoidance of accurate imaging-system models are its main advantage, but the mechanical modulation and multiple-image processing raise the cost and lower the speed. DFF methods also have the requirement that the object shouldn't move during the picture taking, which further limits their potential applications.

The range-sensing technique we proposed also relies on depth from focus, but, unlike conventional techniques, it extracts range with a single unmodulated camera in real-time from just one image.

B. Specific Aims

We have derivated many active and passive range finders that are costly, bulky and slowly. Let us compare them with a fly. 1 μW power runs the brain of a fly, which

weighs less than 1 mg [15]. Of course, a fly cannot tell the range of a tree as precise as $5m \pm 1cm$. Moreover, the bit-error rate of its brain is far from zero. Nevertheless, who has ever seen a fly experiencing difficulties to avoid obstacles? A fly's brain is a special-purpose analog device, designed to deal in real time with imprecise sensory input. This simple creature shows us there are many simple and reliable analog ways to realize the work that usually is done by powerful digital processors.

The purpose of this research is to find a reliable technique that can extract range information at a low cost. The range extraction algorithm is memoryless and simple enough to be implemented on the same CMOS chip integrated with photosensors. The approach presented in this work is a vision system implemented as an optimization problem and realized through a neural network. The choice has been driven by the possiblity of processing the range information locally and directly in hardware, satisfying the real-time constraints [16].

To eliminate the cost of the transmitter and avoid interference of laser or sonar, our technique is based on the passive range finding. We also managed to remove the cost of sophisticate mechanic controllers and digital processors. Since all costly features of vision-based range sensing have been removed, we believe that our technique attains an acceptable performance to cost trade off. As in all vision-based range sensors, however, our technique will be suitable for medium-to-short operation. Therefore, it is not intended as a replacement for long-range sensors needed in most general forward collision avoidance systems, but rather as a specialized component of a sensor-fusion system.

C. A Guided Tour

I will give here a brief preview of each chapter, stating novel results and summarizing the content.

Chapter II reviews previous work on range finders. First, I discuss the category of range finders and list the specs and principles of some commercial range finders. After that, I discuss the principles of different kinds of active ranger finders. Most of the following content is to discuss the concept, theory, advantages and disadvantages of the passive vision based range finders. After addressing Depth from Focus (DFF), we discuss the Depth from Defocus (DFD). This review will help reader to become familiar with the background of range sensing and to understand the advantages of our technique.

In Chapter III, the principle of the proposed technique is discussed. Most of the content involves analysis of the spatial frequency domain response of the system. After giving some background knowledge of optics and image system, we propose our range sensing technique. This system is equivalent to a 4-stage spatial domain filter array. The first filter is the lens used in the system, and the second filter is due to the sampling of image pixels. The third filter array is sensor lines in the tilted chip plane. From the classical Gaussian point spread equation in the vertical image plane, we derive the mathematical model of point spread formation on the tilted plane and use it to analyze the frequency response of the tilted plane. The fourth stage is a high-pass filter or Laplacian filter. Its digital implementation is briefly mentioned. The correctness and feasibility of the technique is verified using a software simulator and hardware emulator. The optical design issues are also discussed.

Chapter IV discusses the architecture of image sensor used in the system. It starts from the fundamental knowledge of photometric, then discusses the physical

principle of photo sensing. After that, I review the background of commercial CCD and CMOS image sensors. After discussing the advantages and disadvantages of the traditional CCD and CMOS active pixels, I introduce the image pixels used in our system. I mainly discuss the approach to widen the dynamic range of the image sensor. Its noise, temporal response, and offset factors are also discussed.

Chapter V is about the circuit architecture that extracts the range information from the frame captured by the sensor in chapter 4. First, we discuss the mathematical principle of the focus measure processor, then discuss how to realize it using analog circuits. The focus measure processor, WTA block, clock generator and readout blocks are carefully analyzed and characterized. Finally, the experimental results of the first prototype system along with the testing environmental setup are provided.

Chapter VI is a very brief conclusion of this work, summarizing the general lesson from this work and proposing future possible work in this system.

I have received help from others especially my advisor in this work, and I want to involve them in the thesis. Therefore, I will use “we” instead of “I” in the dissertation.

CHAPTER II

REVIEW OF RANGE FINDING TECHNIQUES

Range information is the basis of automobile collision avoidance and many other robotic and automatic applications, such as object identification, position tracking, obstacle avoidance, collision avoidance, automation control, human-machine interface utilities, and so on. As we have mentioned in Chapter I, contact distance measurements are not suitable for the collision avoidance. From now on, we only discuss the noncontact ranging techniques.

A. Principle and Category of Range Finding Techniques

The basic principles used in all range finders are [17]:

- Energy propagates at a known, finite, speed.
- Energy propagates in straight lines through a homogeneous medium.
- Energy fields change in a continuous and predictable manner with distance from their source.

Depending on whether transmitters are used, range finding techniques can be divided into the following categories:

Active techniques(wave-based): Involve some form of controlled energy (field or wave) linking a known reference location to the unknown target location.

Passive techniques (field-based): Rely on an externally occurring source of energy (e.g., sunlight or target/background temperature contrast) to make the target detectable. They make use of the spatially distributed nature of an energy form.

The major distinction between field-based approaches and wave-based approaches is that the former, although they employ energy fields, do not rely on the propagation

and conversion (and concomitant losses) of energy. That is, they may employ stationary fields, like those generated by a magnet or static charge. Such fields encode position information by their very shape. Sound and light, although having a wave nature, can be exploited in the same manner as stationary fields because of their distance dependent intensity [17].

Field-based techniques must confront some basic issues that limit their range of application. First, the characteristics of most practically exploitable fields are typically influenced by objects or materials in the vicinity, and it is not always possible to ensure that these influences will remain constant. Second, the variation of fields through space is highly nonlinear (typically inverse square), implying that the sensitivity of a measurement is strongly affected by proximity to the source.

B. Active Range Finding Techniques

Table I lists the specs, principles, and features of some range finders in the market.

From Table I, we can see that most commercial range finders use active techniques. Active range finding techniques greatly simplify the distance measurement problems because they allow a greater degree of control over the many factors. They can always increase intensity of the transmitted signal or change the signal pattern to achieve good signal noise ratio (SNR).

The most commonly used energy forms are: radar, IR laser and ultrasonic. Active range finding techniques are mostly based on time of flight. There are two kinds of techniques based on time of flight: Noncoherent and coherent time of flight.

Table I. Principle and feature of some range finders in the market

Class	Trade Name	Principle	Features
Noncontact (laser)	LASERVISION	TOF	$50m$ range, $4.9mm$ accuracy @$15m$
Noncontact (laser)	HYSCAN	Active triangulation	$40mm$ depth of field, $0.025mm$ accuracy
Noncontact (laser)	ALTM 1020	TOF time-interval	$330 - 1000m$ range, $15cm$ accuracy
Noncontact (laser)	TriCam	Active triangulation	$120mm$ depth of field, $0.05mm$ accuracy

1. Noncoherent TOF

Noncoherent techniques use pulse-bean and measures the time of flight directly. Fig. 1 shows the diagram of range finder using direct TOF. It is especially useful in long-range distance measurements (up to many miles). The transmitter emits very brief, very intense pulses of light. The distance between transmitter and receiver is obtained by counting the time elapsed from the moment when a short train of waves is transmitted to the moment when it arrives at the receiver [18].

$$range(u) = \frac{v \times t}{2} \tag{2.1}$$

The speed and accuracy of these sensors is typically limited by the accuracy with which the time interval can be measured, and the rise time of the laser pulse. In addition, the resolution of the TOF method depends on the ability of generating short-duration pulses and measuring delayed time. For example, in application of

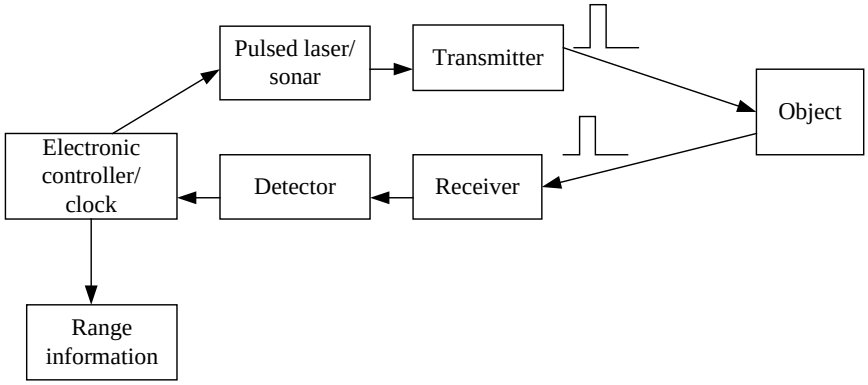


Fig. 1. Block diagram of range finder using direct TOF.

collision avoidance, the typical range is in tens of feet, and the time interval of laser beam is just about a few nanoseconds.

2. Coherent TOF

Coherent time of flight technique uses modulated-beam. This type also uses the time light takes to travel to the target and back, but the time for a single round-trip is indirectly measured by comparing the transmitted signal vs. returning signal. Instead, the strength of the laser is rapidly varied to produce a signal that changes over time. The time delay is indirectly measured by comparing the signal from the laser with the delayed signal returning from the target. One common example of this approach is "phase measurement" as shown in Fig. 2; the phase of the outgoing signal is to be compared with that of the reflected light.

In a coherent range-finding device, both the transmitter and the receiver is fixed on a reference frame. If the distance between the range finder and the object is u and the phase-shift between the transmitted and received signals is $\Delta\Phi$, we have

$$u = \frac{\Delta\Phi v}{4\pi f_c} \tag{2.2}$$

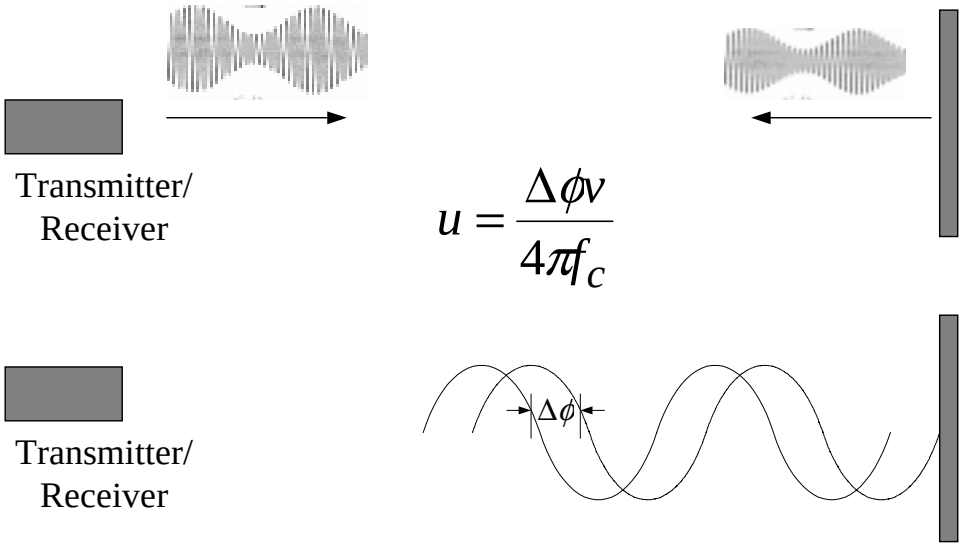


Fig. 2. Range finding using amplitude modulated signal.

where v and f_c are the velocity and frequency of the transmitted signal, respectively.

Phase measurement is limited in accuracy by the frequency of modulation and the ability to resolve the phase difference between the signals.

The frequency-modulated carrier is also frequently used in range finding. Given in Fig. 3 is the principle of frequency modulated TOF. Compared with the amplitude-modulated signal, the range is proportional to beat frequency produced when return is mixed with reference. It has comparatively high noise immunity.

Coherent TOF's superiority over the noncoherent TOF method is obvious: Its dynamic data update rate does not decrease with the increase of the distance between the transmitter and the receiver if no swift variations on the obstacle positions are present. Coherent data output can be easily achieved in real-time multi-channel applications by employing different carrier frequencies for different channels.

For ranges on the order of several feet or less, the total travel time of the light is on the order of about several nanoseconds. So that the travel time of the light is measured to within about one nanosecond or less. Unfortunately, photo-sensitive

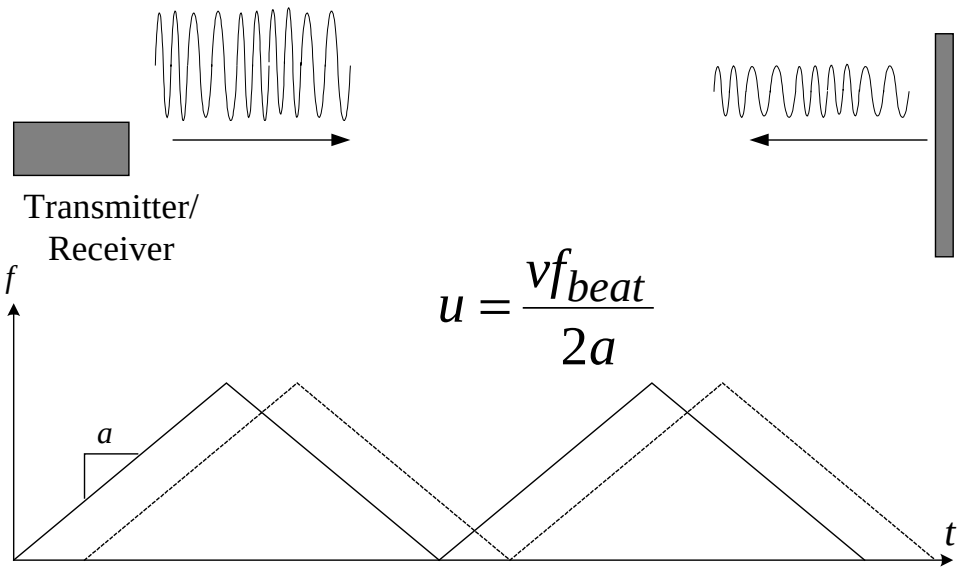


Fig. 3. Range finding using frequency modulated signal.

pixels, such as CMOS active pixel sensors (APS), may have inherent speed limitations on the order of about 5-10 ns. Therefore, usually sonar is used in short distance measuring. But one inherent drawback of using sonar is that the standard digital CMOS technology is insensitive to sonar signals. Therefore it's difficult to integrate the receiver on the silicon.

3. Active Triangulation

Triangulation technique was well known by the ancients. Active triangulation technique also uses transmitters. For distances of a few inches with high accuracy requirements, "triangulation" approach is frequently used. The beam is viewed from one side so the apparent location of the spot changes with the distance of the target. Fig. 4 shows the principles of active triangulation techniques. Lines of detection extend from the ends of the range finder to the target point. If the angles between these lines and distance of baseline can be determined, the distance u is calculated

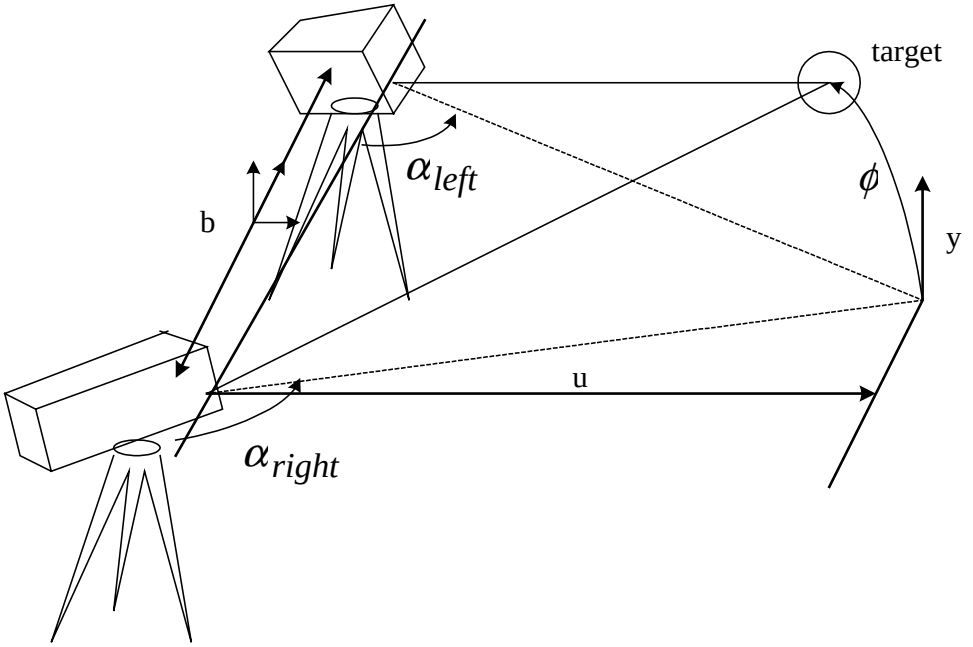


Fig. 4. The basic triangulation geometry as used in classical range finding. The point angles α_{left} and α_{right} are measured locally.

as (2.3) [17]:

$$u = \frac{b \sin \alpha_{left} \sin \alpha_{right}}{\sin(\alpha_{left} - \alpha_{right})} \tag{2.3}$$

Triangulation devices may be built on any scale, but the accuracy falls off rapidly with increasing range, and the depth of field (minimum to maximum measurable distance) is typically limited, as these sensors cannot measure relative to their baseline (distance between emitter and detector).

Whereas TOF and active triangulation techniques employ the wave propagation phenomena of a particular energy form, field-based approaches make use of the spatially distributed nature of an energy form. The intensity of any energy field changes as a function of distance from its source. If the location of a field generator is known and the spatial characteristics of the field that it produces are predictable, remote field measurements contain information that may be used to infer distance from the

source.

Until now, we have investigated the active range finding techniques. We can see that they are wave-based techniques. The following section will discuss the passive techniques. Almost all the passive range finding techniques are field-based.

C. Passive Range Finding Techniques

Depth of field refers to the interval of distance through which a stationary reference ranging system can measure without resorting to a change in configuration. Large depth of field is often an important characteristic in practical applications. For example, if the distance to the target is poorly known a priori, then a large depth of field is desirable.

1. Passive Triangulation

The passive triangulation technique is also called stereovision. It is also based on (2.3), but it doesn't use transmitters. In Fig. 4, each camera in the ends of the base line captures a picture. Assuming epipolar geometry, the problem of reconstructing depth information from two images is very simple if one is able to find the conjugate points in the pair of images.

The difficult task now, of course, is finding the conjugate point in the right image for each point in the left image of a stereogram. This is not an easy job and it is usually processed by powerful digital processor. The following example [10] shows how slow this process can be.

A video-based range finding system would involve at least three components: stereo based object detection, template-based shape matching, and texture-based pattern classification. Assume that each component's performance is independent of

that of the others. It was conservatively estimated that, to detect every object in urban traffic, the stereo component produces one object region of interest (ROI) each 10 seconds, which is too slow for real time range detecting applications.

Compared with active triangulation approaches that have excellent depth of field, passive optical triangulation approaches like those stereography and photogrammetry tend to have restricted depth of field, because the latter rely on camera-type imaging, which is inherently limited by depth of focus.

2. Depth from Focus

a. Background

The depth from focus (DFF) is the most straightforward technique to extract the range information. Many publications [19, 20, 21, 22, 23, 24] have discussed this DFF technique. This category of techniques is based on the thin-lens law as illustrated in Fig. 5.

The DFF methods are based on the following fact: Using an ideal aberration free lens, (1) the focused image faithfully copies the spatial radius distribution of the object surface. The radiance at a point in the scene is proportional to the irradiance at its focused image, and (2) the position of the point of the focus on the object plane can be related to the corresponding point on the image plane by the lens law:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \tag{2.4}$$

where u is the distance from the object plane to lens plane, f is the focal length of the lens, and v is the distance from the focused image plane to the lens plane (see Fig. 5). Given a known lens parameter f, if image distance v can be found, we can easily derive object distance u using (2.4).

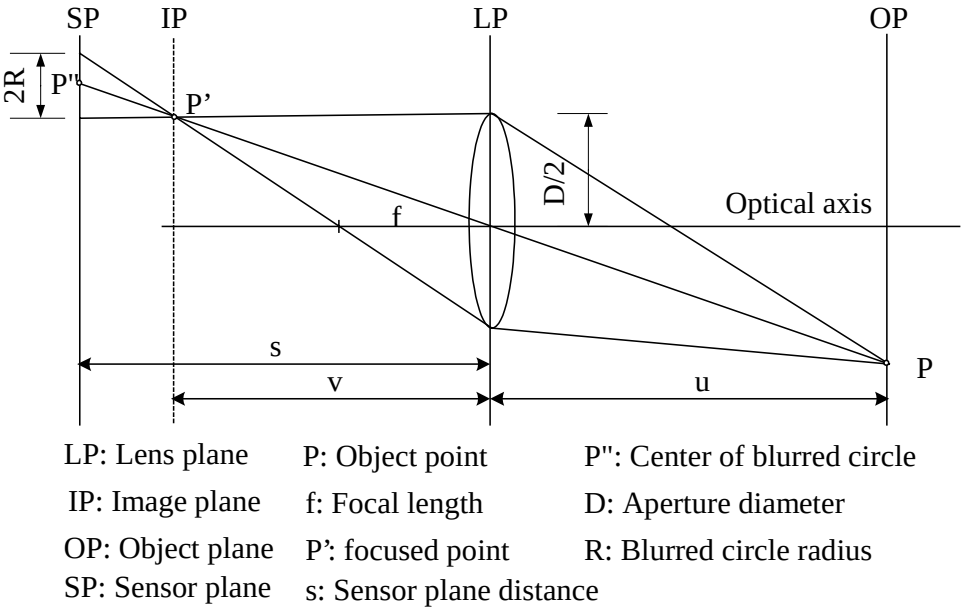


Fig. 5. Image formation using a thin lens.

Fig. 6 [23] shows the focus measure with different lens positions. The sequence of images is obtained by continuously varying the distance between the lens and image sensor. The accuracy of the range information depends on the step size of the lens-sensor distance varying. The basic procedure of DFF is like following: (1) Adjust lens position to acquire multiple images and store them in the memory. (2) Process each saved image and get its focus measure. (3) Pick up the most focused image and use the distance from lens to this image as the image distance v. (4) Use (2.4) to get the distance.

An important advantage of DFF is that, unlike the stereo, it doesn't have the correspondence problem. As a result, it is computationally efficient compared with stereo. Most computation in DFF is to get the focus measure of the image frames. DFF methods use the sharpness of the texture as the focus measure. Fig. 7 shows the images of a pair of shoes captured by moving sensor position from far close to far away from the lens. Among the four photos in Fig. 7, Fig. 7(c) is the most focused

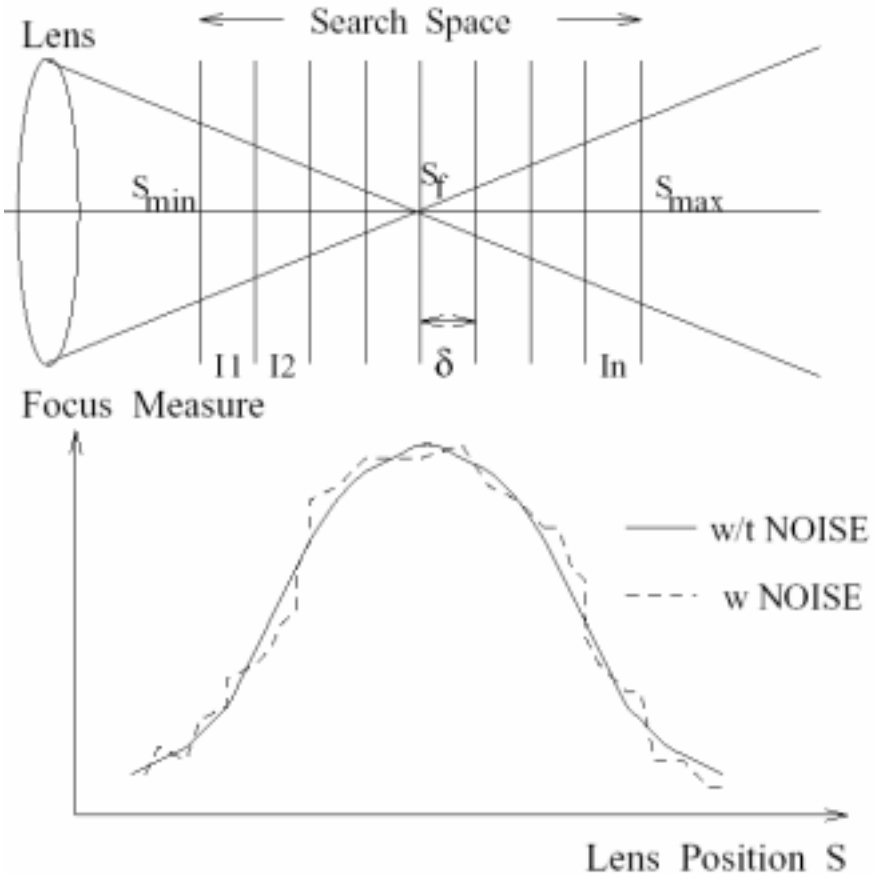


Fig. 6. Image focus analysis with different lens positions.

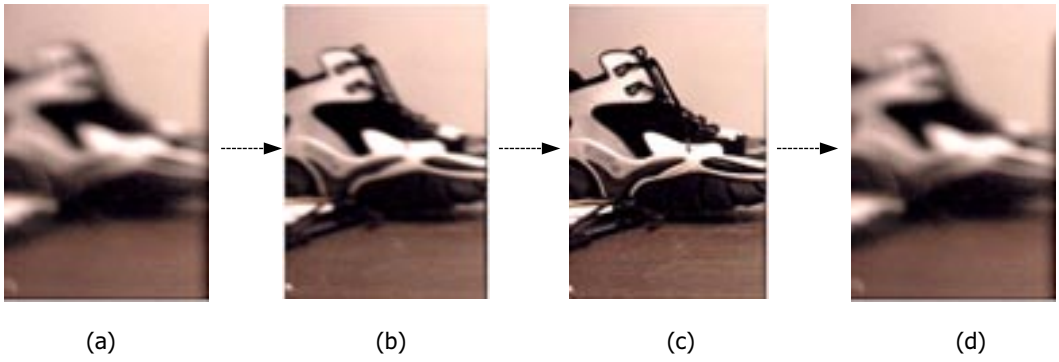
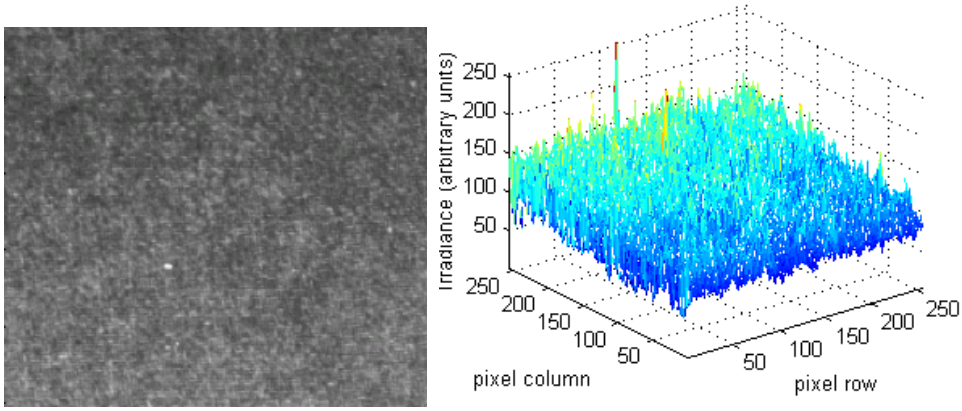


Fig. 7. Four pictures taken with lens-to-film position varying from far close to far away. (c) is the most focused image.

one. Based on the lens parameter of Fig. 7(c), we can extract range information.

Since defocusing is equivalent to low-pass filtering, DFF is effective only if the scene has high frequency brightness variation. It is unsuitable for the objects whose surfaces have no brightness variation. Fortunately, most objects surface has texture; even those surface that appear smooth and non-textured to the naked eye produced high textured images under a good lens. Examples of such surfaces are books, clothes, plastic, walls, etc. Given in Fig. 8(a) is the rough surface of a demin cloth taken from 5 meters away. Although human's eyes can't directly see many textures, the irradiance extraction of Fig. 8(a) shows that the sensor pixels are much more sensitive and can detect many textures (See Fig. 8(b)). Those surfaces are defined as visibly rough surfaces [19]. A surface is considered rough if the dimensions of its spatial variations are comparable to the pixels of the sensor used to observe the surface.

A general focus measure covered by most of the focus measures that have been done by researchers so far is modeled as follows. For each image in the sequence, a focus measure is computed at each pixel in the image. The image for which the focus measure is to be computed is convolved with a focus measure filter. Then the energy of the filtered image is calculated as the image focus measure (See Chapter III for



(a) (b)

Fig. 8. (a) Photo of the rough surface of a demin cloth. (b) The extraction of its surface irradiance using a Matlab program.

more details).

In all the DFF techniques proposed until present, there are some common constraints:

- A large number of images is needed as inputs to compute a focus measure in order to determine the focused image. To capture a large number of images, it needs mechanical motion of camera parts to change lens position in order to acquire images and hence quite slow. Note that in order to get a fine range resolution, the step size of the mechanical adjustment must be very small.

- A/D converters and memory are needed to convert and store photo voltage of sensor to digital signal for further processing, which is done by a separate digital processor.

3. Depth from Defocus

Some researchers proposed the range finding methods that don't need the focusing information of the object [13, 14, 25, 26, 27, 28]. They took the level of defocus of

the image and corresponding camera parameter values into account in determining distance. This approach is called Depth from Defocus (DFD). Pentland [13, 14] noticed the fact that most biological lens systems are exactly focused at only one distance from the lens to the scene. As the distance between the imaged point and the surface of exact focus increase or decreases, the imaged objects become progressively more defocused.

Pentland [13] showed that distance u to an imaged point is related to the parameters of the lens system and the amount of defocus by the following equation (see Fig. 9):

$$u = \frac{fs}{s - f - 2R_1N} \tag{2.5}$$

$$u = \frac{fs}{s - f + 2R_2N} \tag{2.6}$$

for $sp < v$, where N is the f-number of the lens ($N = \frac{f}{D}$). It is clear that a single image does not include sufficient information for depth estimation as two scenes defocused can produce identical images. Therefore, two images are needed to achieve depth information. As far as arbitrary scenes are concerned, Pentland proposed another algorithm that uses Fourier transform to find the blur circle radius.

a. Fourier Approach

This section briefly discusses how to use Fourier approach to get depth information [11]. The image formation in a camera with variable lens parameters (u, s, f) is shown in Fig. 5. According to geometric optics, the normalized radius of the blur circle is

$$R = \frac{D}{2}(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}) \tag{2.7}$$

Accurately enough, the Point Spread Function (PSF) of the lens can be repre-

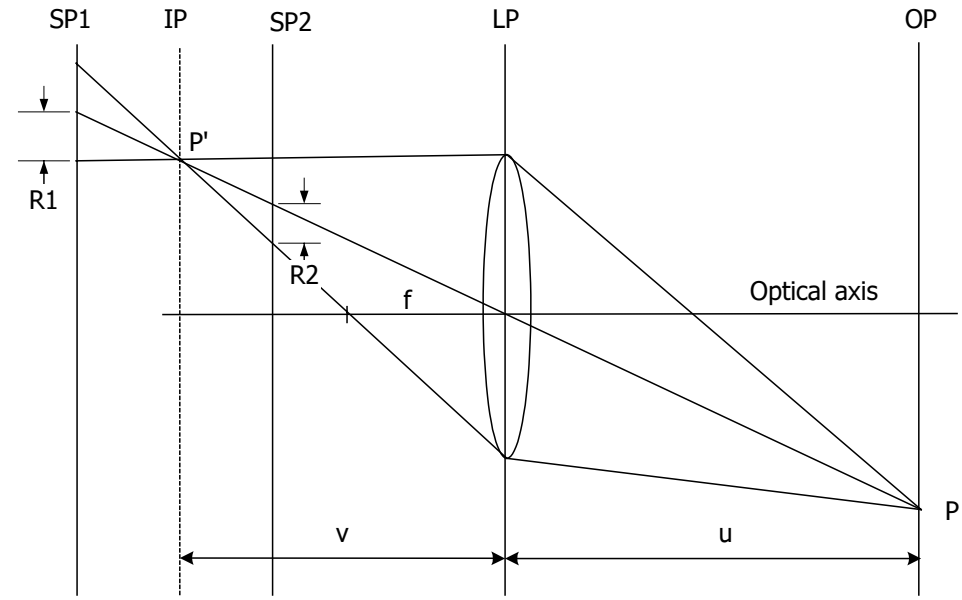


Fig. 9. Geometry of defocused images on different sensor planes.

sented by a two-dimensional Gaussian:

$$h(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2.8}$$

where the spread parameter σ is proportional to the blur circle radius R ($\sigma = kR$), where k is a lens constant [19]. A blurred image g of a planar object having focused image f is given by the convolution h of the PSF and the focused image f, $g = f * h$. In the frequency domain, the relation becomes $G(w,v) = H(w,v)F(w,v)$ where $H(w,v) = e^{-\frac{1}{2}(w^2+v^2)\sigma^2}$. For two blurred pictures g_1, g_2 taken with two different camera setting, we have $G_1(w,v)/G_2(w,v) = e^{-\frac{w^2+v^2}{\sigma_1^2-\sigma_2^2}}$, or

$$\sigma_1^2 - \sigma_2^2 = \frac{-2}{w^2+v^2} \ln \frac{|G_1(w,v)|}{|G_2(w,v)|} \tag{2.9}$$

Then from (2.7), we have

$$\sigma_1 = k\frac{D}{2}(\frac{1}{f} - \frac{1}{u} - \frac{1}{s_1}),\ \sigma_2 = k\frac{D}{2}(\frac{1}{f} - \frac{1}{u} - \frac{1}{s_2}) \tag{2.10}$$

Eliminating $1/u$ from above equations we obtain

$$\sigma_1 = \alpha\sigma_2 + \beta, \text{ where } \alpha = \frac{D_1}{D_2},, \beta = kD_1(\frac{1}{s_2} - \frac{1}{s_1}) \tag{2.11}$$

From (2.9) and (2.10), we have

$$(\alpha^2 - 1)\sigma_2 + 2\alpha\sigma_2 + \beta^2 = \frac{-2}{w^2+v^2} \ln \frac{|G_1(w,v)|}{|G_2(w,v)|} \tag{2.12}$$

Since σ_2 is the only unknown variable in (2.12), we can solve it. Putting the value of σ_2 back to (2.10), we can get distance u.

The above discussion briefly discusses the feasibility of DFD. To get the image depth, the above procedure needs to be repeated for the entire image neighborhood. Therefore, the calculation is quite heavy. More details of DFD algorithm can be found in[29].

The advantage of DFD is that it does not involve searching for image distance v from very many frames, which is used in (2.4) to get the object distance u. Therefore, these methods require processing only a few images (about 2-3) as compared to a large number of images in the Depth from Focus (DFF) methods.

Compared to DFF, some disadvantages of the DFD compared to DFF are as follows: (1) DFD requires accurate camera calibration for the camera characteristics (point spread function as a function of different camera parameters). (2) DFD is less accurate than DFF methods. (3) DFD data processing is much heavier and takes longer time.

D. Investigation Conclusions

All vision-based (field-based) techniques discussed so far have two major constraints:

- They need more than one image to extract range information. Mechanical adjustment is unavoidable.
- They need to store images in memory so that the image can be processed later by a digital processor. So A/D converter, memory, and digital processor are imperative.

In the next chapter, we will propose our technique to overcome those two constraints.

CHAPTER III

PRINCIPLE OF THE PROPOSED SYSTEM

In this chapter we propose a range finding technique and discuss its design optimization. Before discussing its principle, we briefly introduce image formation and imaging systems as a prerequisite for full understanding of how to recover range information from the image.

A. Imaging System

A typical imaging system is shown in Fig. 10. The goal of imaging system is to provide sufficient image quality to enable extraction of desired information about the object from the image. Some of the components of imaging quality are resolution, image contrast, perspective errors, geometric errors (such as distortion) and depth of field. First we briefly introduce the terminology used in imaging system.

1. Energy, Flux and Intensity Measurements

There are two types of terminologies in imaging system: radiometric and photometric. For example: Light source can be measured by radiance ($W/sr/m^2$) or luminance (cd/m^2); image can be measured by irradiance (W/m^2) or illuminance (lux). Radiometry is concerned with the measurement of electromagnetic energy. It is purely physical. How the (standard) human eye records optical radiation is often more relevant than the absolute physical values. This evaluation is described in photometric units and is limited to the small part of the spectrum called the visible. Photon quantities are important for many physical processes. Photometry takes into account the visual effectiveness of the light to the sensitivity response of human eye. Photometric measurements are based on the photonic-eye response unless otherwise stated [1].

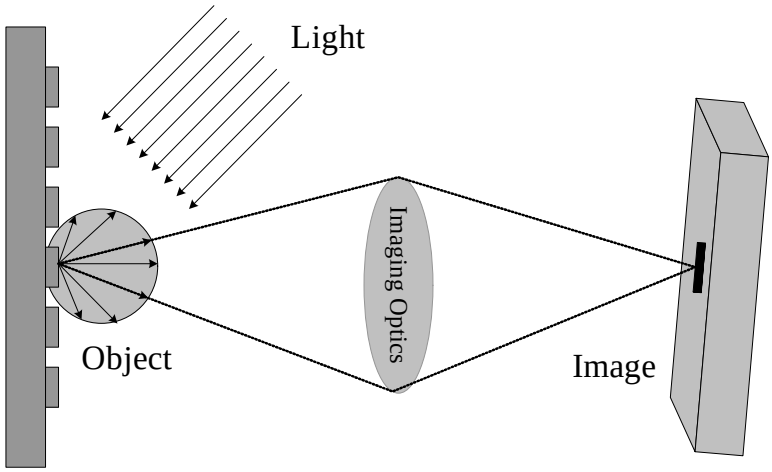


Fig. 10. Typical imaging system using a nonideal lens.

2. Terminology

The terminology associated with electromagnetic-radiation measurement is shown in Table II.

Radiometric terms can be converted to the matching photometric quantity. The photometric measure depends on how the source appears to the human eye. This means that the variation of eye response with wavelength, and the spectrum of the radiation, determines the photometric value. Invisible sources have no luminance, so a very intense ultraviolet or infrared source registers no reading on a photometer.

Conversion from a radiometric quantity (in watts) to the corresponding photometric quantity (in lumens) simply requires multiplying the spectral distribution curve by the photonic response curve, integrating the product curve and multiplying the result by a conversion factor of 683. Mathematically, the relation between a photometric quantity (PQ) and its matching radiometric quantity (SPQ) is given by:

$$PQ = 683 \int (SPQ_\lambda) V(\lambda) d\lambda \tag{3.1}$$

Table II. Radiometric and photometric terms and equations [1]. Symbols Key:J: joule, lm: lumen,W: watts, s: second,m: meter, cd: candela,sr : steradian, lx: lux, lumen m^2

Radiometric	Photometric	Equation	Note
Radiant energy (J)	Luminous energy (lms)	Q	Energy
Radiant flux (W)	Luminous flux (lm)	$\phi = \frac{dQ}{dt}$	Power, flux
Irradiance (W/m^2)	Illuminance $(lm/m2)$	$E = \frac{d\phi}{dA}$	Power output per unit area
Radiant exitance (W/m^2)	Luminous intensity (lm/m^2)	$M = \frac{d\phi}{dA}$	Power per unit area
Radiant intensity (W/sr)	Luminous intensity (cd)	$I = \frac{d\phi}{d\Omega}$	Power per unit solid angle
Radiance (W/m^2sr)	Luminance (cd/m^2)	$L = \frac{dl}{dA\cos\theta}$	Power per unit solid angle per unit projected

Table III. Illuminance table

Direct sunlight	100,000-130,000 lux
Full daylight, indirect sunlight	10,000-20,000 lux
Overcast day	1,000 lux
Indoor office	200-400 lux
Very dark day	100 lux
Twilight	10 lux
Full moon	0.1 lux
Moonless clear night sky	0.001 lux

3. Illuminance Table

The illuminance of the natural light can have more than 6 decade variance. Table III lists lux number of some typical illumination [30] .

B. Lenses and Image Formation

The dissertation uses the following symbols to specify the imaging system:

u: Object to front principal point distance

v: Rear principal point to image distance

f: Focal length

M: Magnification

D: diameter of the entrance pupil, i.e. diameter of the aperture seen from the front of the lens

N: f-stop (or f-number) $N = f/D$

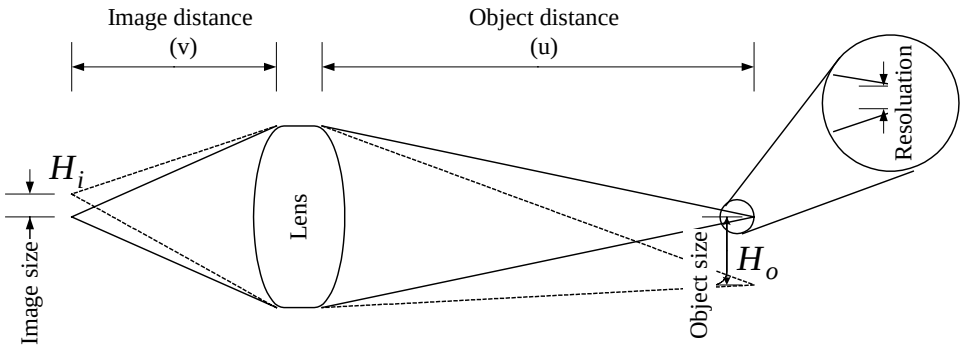


Fig. 11. Image formation through a convex lens.

An image is a two-dimensional pattern of brightness. The lens used in our system is a conjugate lens. The image formation through the lens is shown in Fig. 11. The relationship between the focused image and object is given by:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \tag{3.2}$$

$$M = \frac{H_i}{H_o} = \frac{v}{u} \tag{3.3}$$

Note the lens in Fig. 11 is a thick lens (different from the thin lens in Fig. 5), which is the combination of several simple lenses whose individual optical axes are carefully lined up. The reason to avoid simple lens is that simple lens has a number of defects or aberrations [31].

a. Irradiance of Imaging

The irradiance of the image depends on the object radiance, lens system, and sensor position. As shown in Fig. 12, flux collected by the lens is given by [32]

$$d\Phi = \int_0^{2\pi} d\varphi \int_0^{\theta_0} L dA_0 \cos\theta \sin\theta \, d\theta \; [W]$$

Usually when the sensor is on-axis, the image resolution will be higher. The

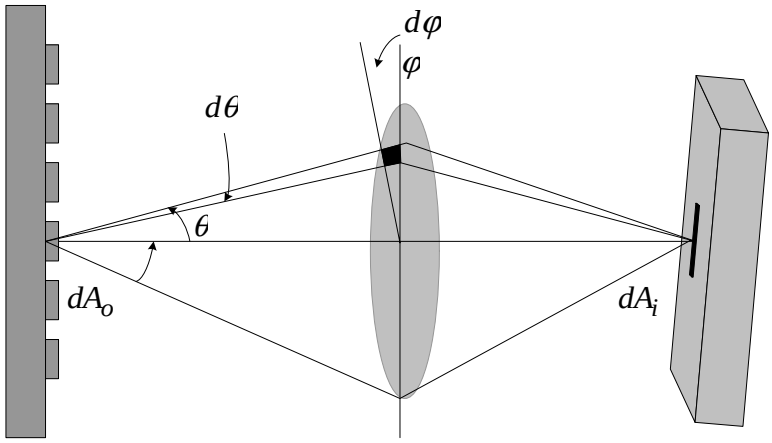


Fig. 12. On-axis flux collection.

on-axis image irradiance is given by

$$E = \frac{d\Phi}{dA_i} \cong \frac{\pi}{4}\frac{1}{N^2(1-M)^2}L \ [W/m^2] \tag{3.4}$$

Of course not all the images are on-axis image. Fig. 13 shown the off-axis image, whose irradiance is given by

$$E = \frac{d\Phi}{dA_i} \cong \frac{\pi}{4}\frac{1}{N^2(1-M)^2}\cos^4\phi L \ [W/m^2] \tag{3.5}$$

If the area of sensor plane is large, e.g. $35mm$ film, the effect of off-axis image can't be neglected. In our system, since the area of chip ($2.4mm \times 2.4mm$) is much smaller compared with the focal length, we can assume all the images captured by the sensor are on-axis.

b. Image Resolution and Contrast

Resolution is a measurement of the imaging system's ability to reproduce object detail. Determining the minimum necessary space yields the limiting resolution of

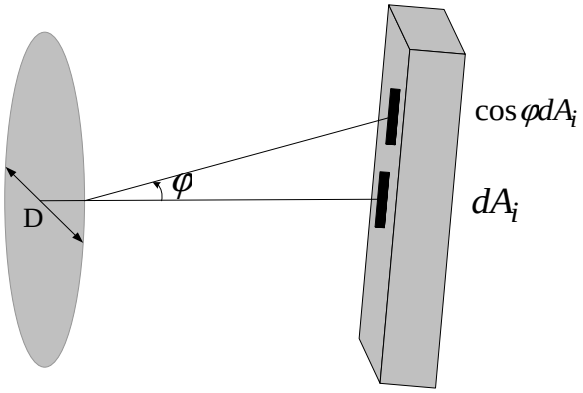


Fig. 13. Off-axis flux collection.

the system. This relationship between alternating black and white squares is often described as a line pair. The resolution is typically defined by the frequency measured in line pairs per millimeter (lp/mm) [33].

Contrast, which describes how effectively the differences between boundary areas on the image are reproduced relative to one another, can often be defined in terms of grayscale or signal-to-noise. For an image to appear well defined, the black details must appear black and the white details, white (see Fig. 14). The greater the difference in intensity between a light and a dark line, the better the contrast. The contrast is the separation in intensity between blacks and whites:

$$\%contrast = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

Resolution and contrast are closely linked. In fact, resolution is often meaningless unless defined at a specific contrast. Similarly, contrast depends on resolution frequency.

After brief introduction of basic knowledge of the imaging system, we will discuss the proposed range system in the next section.

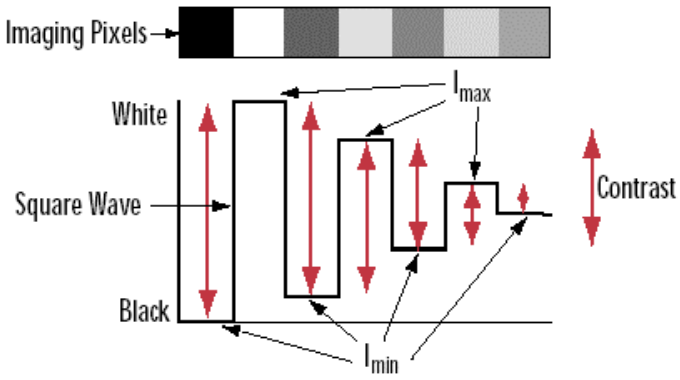


Fig. 14. Contrast is the difference in intensity between blacks and whites. For an image to appear well defined, black details must appear black and white details must appear white. The greater the difference in intensity between a black and white line, the better the contrast. The human eye can see a contrast of as little as $1-2\%$. Our imaging system is designed to see a typical limiting contrast of 10% to 20%.

C. Proposed Technique

Our system is a special Scheimpflug camera. Scheimpflug cameras are distinguished by a sensor plane tilted at a non-orthogonal angle with respect to the optical axis. This unique property leads to a non-frontal focusable-object plane (FOP), as illustrated in Fig. 15. The FOP can be uniquely determined by (1) applying Scheimpflug principle, according to which, the sensor and lens planes and FOP must intersect along the same line, which is identified as "Scheimpflug line" in Fig. 15(a), and (2) the Gaussian Lens Law to find the distance $z = u$ of the point where the optical axis intersects FOP. These cameras are used in photography to capture focused images of tall structures or wide fields and in ophthalmology for cataract detection [34]. They also find use in particle-image velocimetry [35]. To the best of our knowledge, there exists only one range estimation technique utilizing a Scheimpflug camera [36]. It is applied by panning the camera in incremental steps around the axis x, storing the image frame acquired at each step, and comparing the frames for the degree of focus. Since

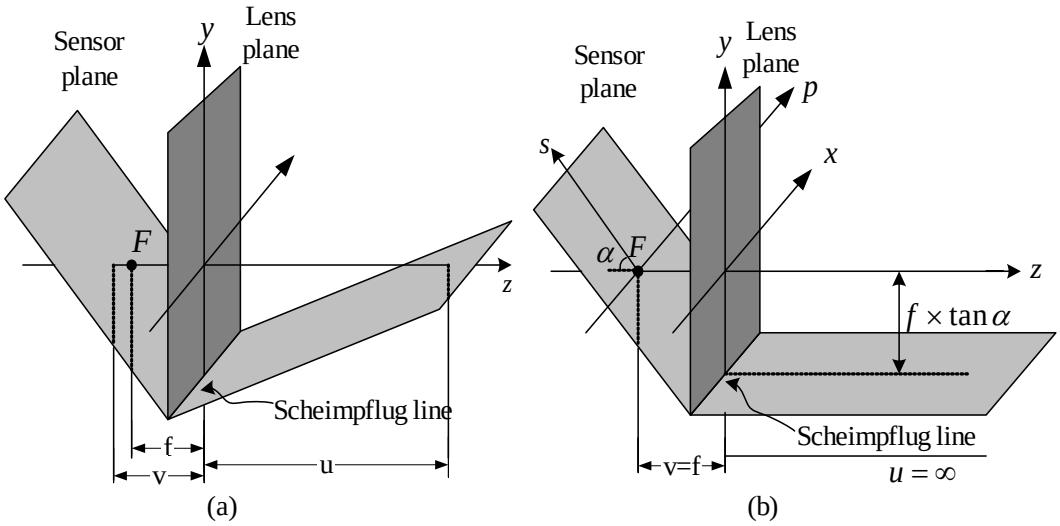


Fig. 15. The sensor, lens and focusable-object planes in a Scheimpflug camera. The Scheimpflug Principle states that all three must intersect along the same "Scheimpflug line." (b) The configuration of the Scheimpflug camera proposed in this work for range sensing. Note the introduction of sensor-plane coordinates p and s, whose origin is at the focal point.

panning causes FOP also to rotate around the axis x, the image of a point object is maximized in the frame where the object point falls onto the FOP. This technique is capable of extracting range maps in two dimensions along x and y, but it needs mechanical panning, multiple frame storage, and complex signal processing algorithms to eliminate frame correspondence problems.

Our range-sensing technique is implemented with the Scheimpflug camera configuration depicted in Fig. 15(b). Note that the sensor plane not only is tilted at a non-orthogonal angle α, but it also intersects the optical axis at the focal point F. The latter condition moves the point of intersection of FOP with the optical axis to infinity, and thus forces the FOP to be oriented parallel to the optical axis with an offset

$$y = -f \tan \alpha \tag{3.6}$$

is determined by the y-coordinate of the Scheimpflug line alone. As depicted with 3-D and 2-D schematics in Fig. 16, objects with different distances will have focused images on the different parts of the sensor. In this simple geometric arrangement, only the objects lying on the lateral plane $y = -f/\tan\alpha$ are focusable on the tilted sensor plane because no parallel ray approaching the lens plane at any other elevation can be traced onto the sensor plane after passing through F. Now consider a planar object intersecting the FOP perpendicularly at a distance z_o, as shown in Fig. 17(a) where the configuration is viewed along x.

The significance of this focusable-object plane (FOP) is illustrated in Fig. 17(a) with points O_1 and O_2 of an object plane intersecting FOP perpendicularly. Although both points are at the same distance to the lens, only the image I_1 of O_1 is in focus because the latter is in FOP. Since the point of focus I_2 of O_2 is below the sensor plane, the image of O_2 on the sensor plane is spread into a conic-sectional area. Naturally, this spreading results in a blurred image well approximated by an ellipse. Any object point located above the FOP projects a similarly blurred image because its point of focused image develops above the sensor plane. The sensor-plane coordinates of the focused image of a point on FOP can be determined from Gaussian lens law as follows:

$$s_i = \frac{f^2}{(z_o - f)\cos\alpha} \cong \frac{f^2}{z_o \cos\alpha} \tag{3.7}$$

$$p_i = \frac{z_o f}{(z_o - f)\tan\theta} \cong f \tan\theta_0 \tag{3.8}$$

Where z_o and θ_0 are, respectively, the range and bearing of the object point. Obviously, the range and bearing of the points lying on FOP can be extracted from

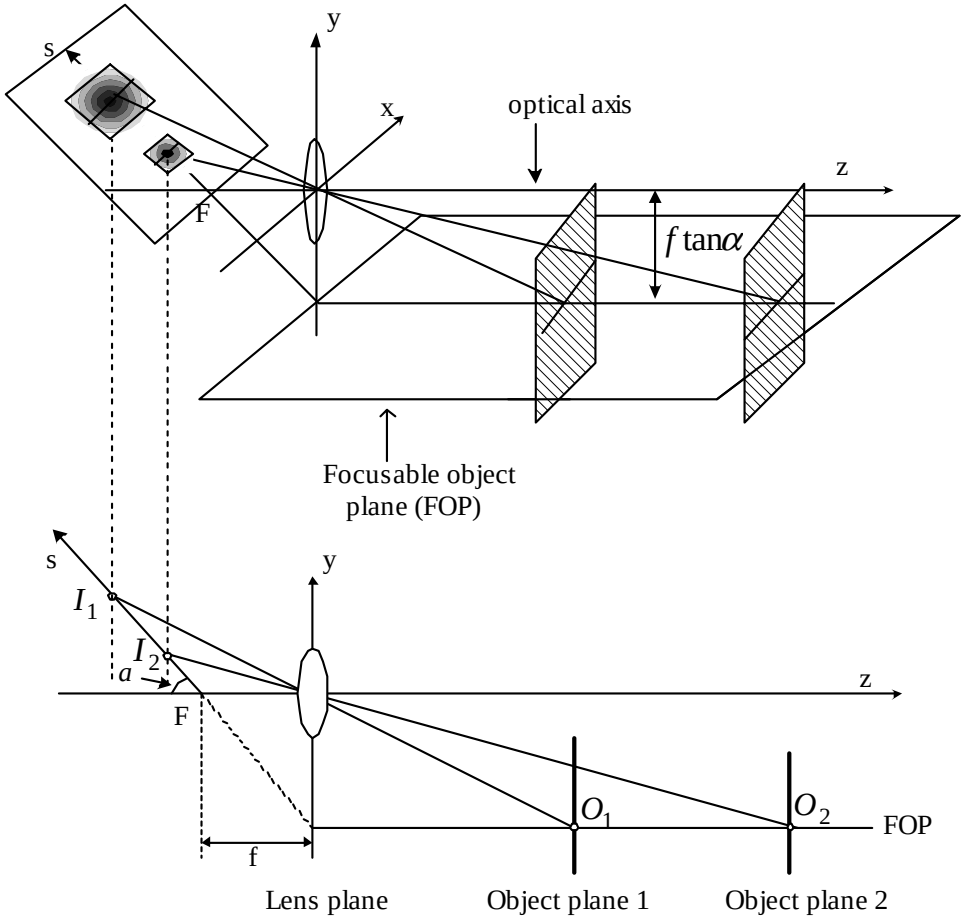


Fig. 16. 3D and 2D system plot of images of two objects with different distances. Note that object plane 1 will occlude object plane 2 if they are range sensed at the same time.

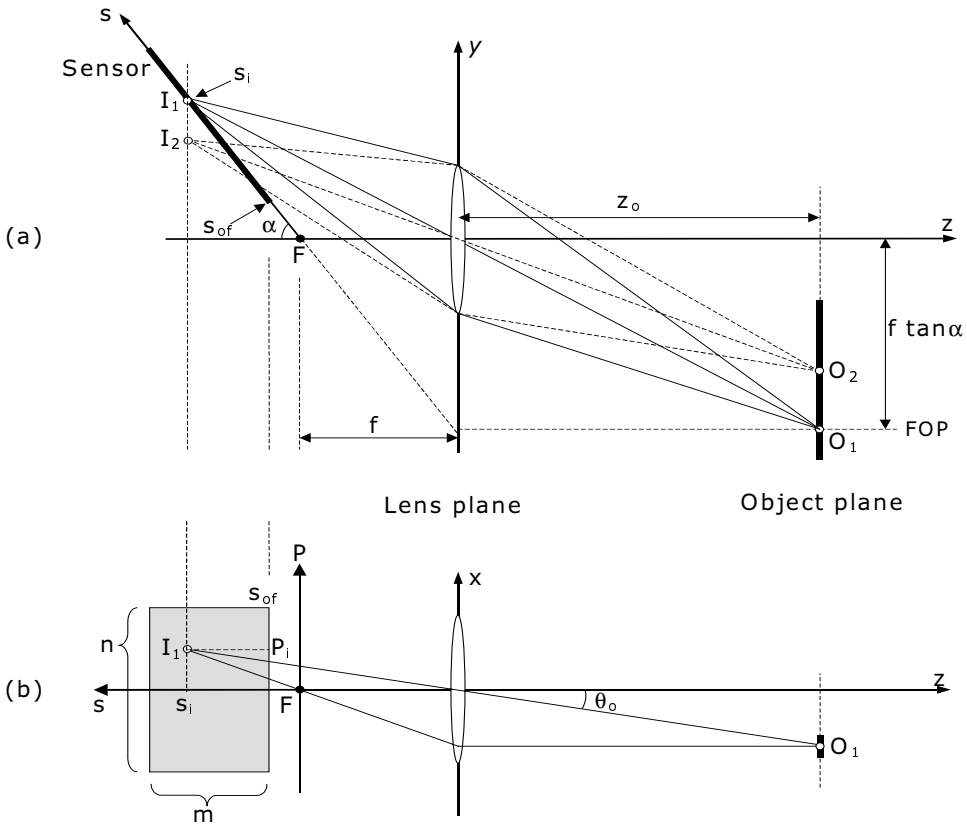


Fig. 17. Formation of focused and defocused images on the sensor plane. Note that the sensor has m rows and n columns of pixels, and is offset from the focal point by s_{of}. (a) View along x. (b) View along y.

(3.7) and (3.8) if only the sensor coordinates of their focused images can be identified. Assuming a pixel matrix of m rows and n columns of pixels as shown in Fig. 17(b) where the configuration is viewed along y, we therefore can resolve the range and bearing of these object points in, respectively, m and n discrete quanta without modulating any lens parameter.

The predictions of (3.7) and (3.8) are perfectly verified with the experiment illustrated in Fig. 18. The image shown in Fig. 18(b) belongs to four objects of $5mm \times 5mm$ checkerboard pattern positioned at different distances perpendicularly to the plane xz as shown to scale in Fig. 18(a). The Scheimpflug camera used for

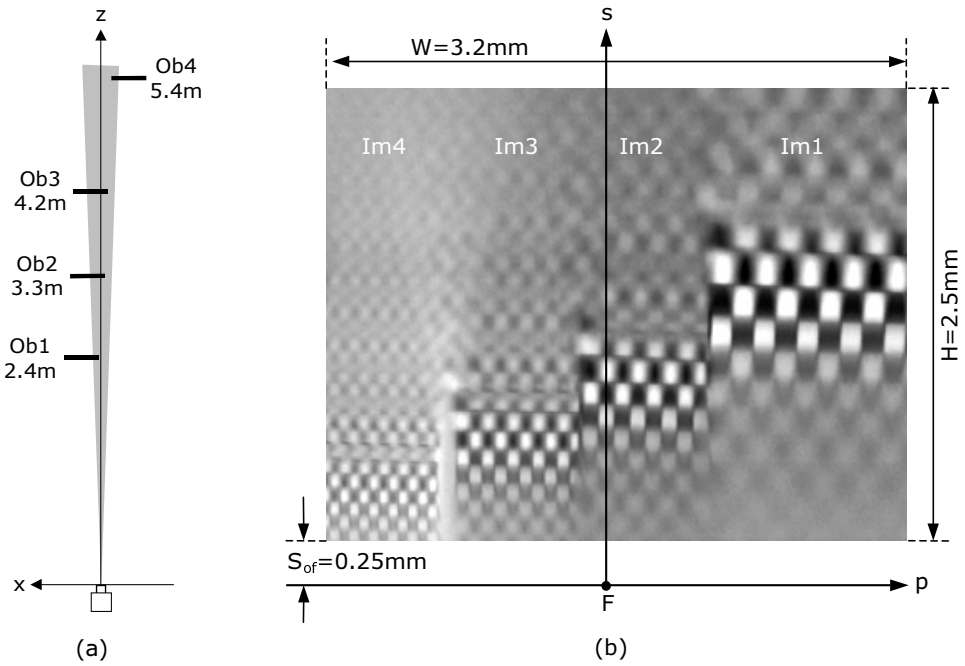


Fig. 18. (a) Experimental setup for ranging four objects at different distances and bearing angles. The shaded triangle is the field-of-view FOV_x on the plane xz. (b) The video image captured with the following Scheimpflug camera settings: $\alpha = 45^o$, $f - stop = 1.4$, $s_{of} = 0.25mm$.

capturing this image was originally built as a test bench for the analytical model we developed for focus roll-off on the tilted sensor plane. It is equipped with an Omnivision OV5116 1/4in video chip of 352×288 pixels whose width and height are specified as $W = 3.2mm$ and $H = 2.5mm$, respectively. The lens is an $f = 50mm$ Pentax SLR. Adjustable camera settings for this particular image were $\alpha = 45^o$, f-stop=1.4, and the chip was offset from F by $s_{of} = 0.25mm$. In collision avoidance applications, we are interested in sensing the range of only the closest obstacle at any given bearing. Our focus detection procedure is highly simplified to facilitate its implementation with an embedded analog network. It is based on two conditions to be satisfied by the object to be ranged. First, we assume that the field-of-view for any sensor column in the direction perpendicular to the plane xz is

fully occupied by the same object. We denote this field-of-view with FOV_y in Fig. 19, where an object satisfying this condition is seen to fully occlude the scene behind it. This occlusion condition rules out the possibility of having focused images of multiple objects projected onto the same sensor column. Therefore, we can determine the coordinate s_i corresponding to the single closest object simply by conducting a competitive maximum-selection process among the focus measures collected by all pixels of the column.

A narrower FOV_y is needed for another but closely related reason. Whatever focus measure is adopted for detection, however, its value detected at a pixel depends not only on whether the local image is in focus but also on the spatial-frequency content of the object pattern. In order for the competitive maximum-selection process to yield the coordinate of best focus without ambiguity, the pattern of the object patch viewed by all the pixels of the corresponding sensor column must remain substantially invariant along y. This pattern-invariance condition is met if the object pattern falling into FOV_y is comprised mostly of edges oriented along y. In our originally targeted application area of highway collision avoidance, most of the objects of collision potential indeed exhibit edges, most notably horizontal ones. By aligning y horizontally, we therefore increase the probability of satisfying the pattern-invariance condition. In any case, a narrower FOV_y generally helps raise the probability of satisfying both conditions. Assuming f to be relatively larger than H, FOV_y is described approximately by

$$FOV_y = \frac{H \sin \alpha}{f} \quad (rad). \tag{3.9}$$

Theoretically, FOV_y can be reduced to zero by setting $\alpha = 0$ and forcing FOP to be aligned with the optical axis but this will also excessively reduce the irradiance.

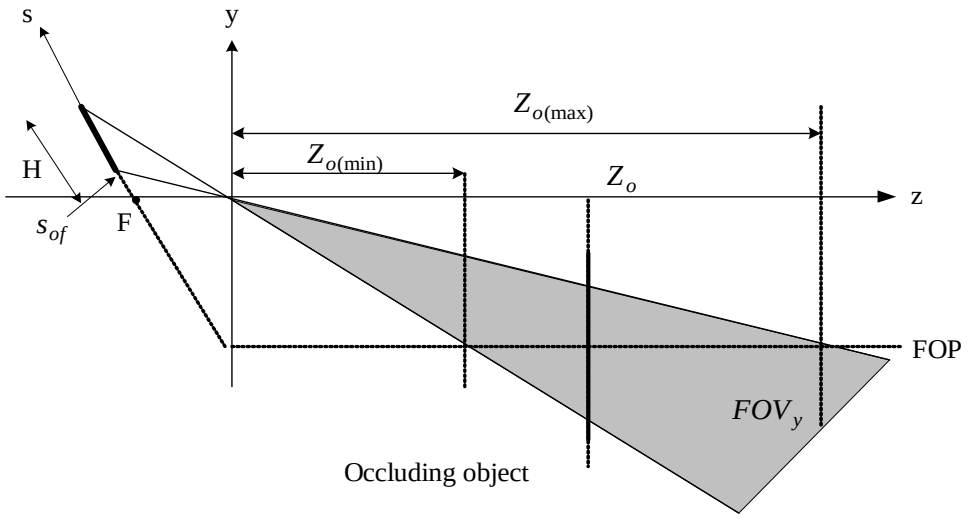


Fig. 19. Occlusion condition imposed on the field-of-view FOV_y in the direction perpendicular to the plane xz. The minimum and maximum limits, $z_{o(min)}$ and $z_{o(max)}$ of the range sensed are determined by s_{of}, H, f and α.

The value of FOV_y in our experimentation of Fig. 19 is 3.5×10^{-2} radians.

A simple trigonometric derivation indicates that the extreme elevation of FOV with respect to FOP is no more than $10cm$ for a lens aperture and f in the centimeter range, and the ratio of maximum range to minimum range on the order of 10. With such a narrow TFOV, the crucial condition of occlusion can be easily satisfied by any obstacle of collision significance. Our technique is not capable of range mapping along y because FOV_y is utilized solely for focus detection as just explained but range mapping along x is possible as already demonstrated in Fig. 19. The field-of-view along this direction, FOV_x, is described by

$$FOV_x = \frac{W}{f}(rad), \tag{3.10}$$

which measures 6.4×10^{-2} radians, and identified with a shaded triangle in Fig. 18(a).

D. System Design

Although camera optimization is closely related to the objectives of the application selected, it is still subject to a number of constraints emanating from the general optical configuration. The minimum and maximum of the range covered and range-resolution are three such strongly performance metrics.

As explained above, our technique extracts the point of best focus by applying a column-wise maximum-selection process on all focus-measure values collected by the pixels of the column. The accuracy of this process increases with (a) the selectivity of the maximum-selection algorithm used, and (b) the rate by which the adopted focus measure rolls off with s around its peak at s_i.

1. Range Limits and Resolution

The minimum and maximum of the range can be determined from Fig. 19. Again assuming $f \ll H$, these limits are approximately described by

$$z_{o(min)} = \frac{f^2}{(s_{of} + H)\cos\alpha}, \tag{3.11}$$

and

$$z_{o(max)} = \frac{f^2}{s_{of}\cos\alpha}. \tag{3.12}$$

The range resolution of the camera, defined as the relative variation in z_o that shifts s_i by one row on the sensor plane, can be calculated from (3.7). Assuming the sensor-row pitch H/m to be negligible in comparison with s_i, the result is as follows

$$\frac{\Delta z_o}{z_o} = \frac{H\cos\alpha}{mf^2} z_o,$$

which clearly shows that the resolution deteriorates with the range z_o as in all image-based range sensing techniques.

Some general design tradeoffs readily emerge from (3.9) and (3.11)-(3.12). First, assuming that the pitch H/m is already minimized for a given technology, and considering that $\cos\alpha$ is already close to unity for a small tilted angle, the resolution can be improved only by selecting a longer focal length, which also help reduce FOV_y and increase the maximum range. Unfortunately, it also increases the minimum range. Attempting to counter this effect by increasing H not only increases chip size but also adversely affects FOV_y. The only remaining camera parameters, s_{of}, is only weakly effective on the minimum range unless it is made comparable to H.

2. Frequency Response of the System

From Fig. 18, we can see that the focus peak position is the criterion to find range, and focus roll-off will affect the resolution of range finding. In order to determine how camera parameters affect focus roll-off and focus peak, one need a quantitative model for image projection onto a tilted sensor plane. Although a numerical model can be developed on the basis of the procedure proposed in [37], an analytical model is more preferable for design purposes. The proposed system is a multistage spatial frequency domain filter array, which is shown in Fig. 20.

In the following sections, we will discuss the frequency response of each stage of filter.

a. Lens Filter

All the energy of image is collected by the lens, whose quality dramatically affects the image quality and the focus peak value. Ideally we would expect that the lens can reproduce image with infinite high resolution and infinite high contrast, but the real-world lenses mostly probably provide picture in Fig. 21.

The nonideality of lens is due to the following reasons:

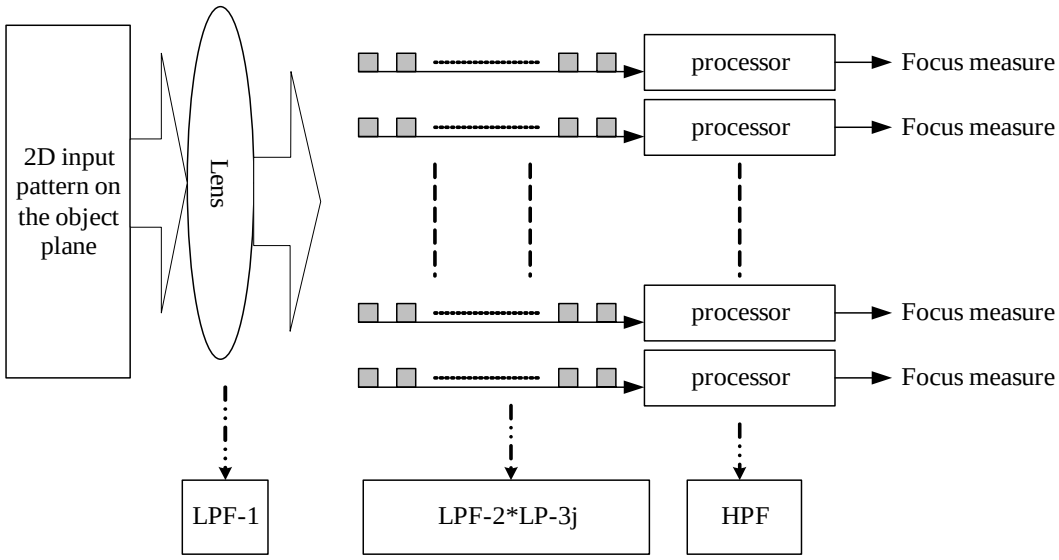


Fig. 20. Range finder as a four-stage spatial domain filter: Lens: LPF-1; sampling sensors: LPF-2; tilted sensor plane: LPF-3j (j is the index of pixel row whose frequency response is LPF-3j); processor, HPF. "*" is the convolution symbol.

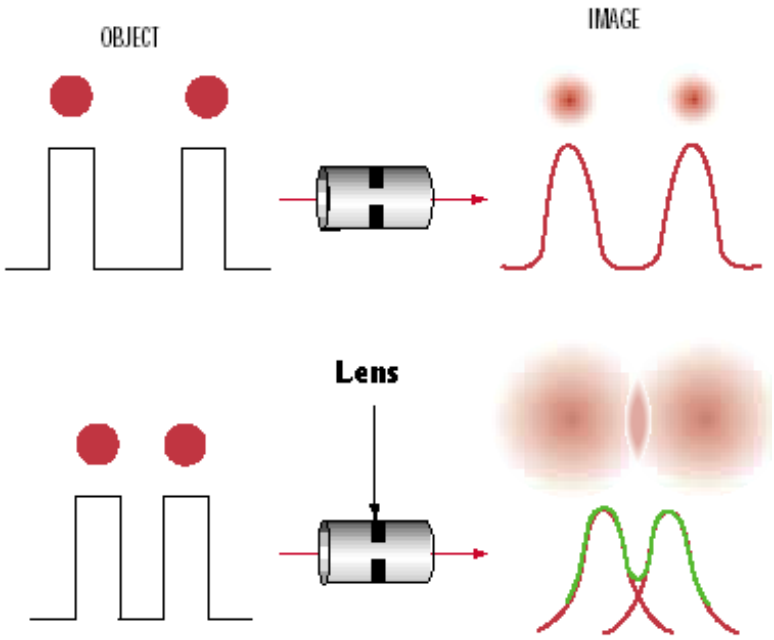


Fig. 21. The output of the image collected by a lens. Note that the high frequency components are attenuated.

Diffraction

The edges of the aperture create a certain amount of fuzziness. This is what is called diffraction. The smaller the aperture, the more the spreading [38].

Aberration

Aberration arises simply because in the real world lenses have finite thicknesses and all rays are not strictly paraxial as we assumed to obtain the Gaussian formula (3.2). Aberration correction is a large part of what distinguishes mediocre from excellent lenses. As we can see in Fig. 21, lenses behave like low pass filter (LPF). The better the quality of lens, the higher stop frequency of the lens. The criterion used by optical industry to characterize the frequency response of the lens is Modulation Transfer Function (MTF).

MTF provides a graph analyzing a lens' ability to resolve sharp details in very fine sets of parallel lines, and a lens' contrast or ability to provide a sharp transfer between light and dark areas in sets of thicker parallel lines. MTF is also a frequency response, except that it involves spatial frequency-cycles (or line pairs) per millimeter instead of per second. High spatial frequencies correspond to rapid changes in image density, i.e., fine image detail. The response of lenses tends to roll off at high spatial frequencies. Therefore, lenses can be thought of as low pass filters-filters that pass low frequencies and attenuate high frequencies.

Let the input to the imaging optics be a $1-D$ sinusoidal monochromatic photon flux $\Phi(x, fi)$ and let the response be a photon flux $\Phi'(x, f_i)$. Fig. 22 shows the response of a typical lens.

MTF of the lens at frequency f_i is defined by:

$$MTF(f_i) = \frac{b'/a'}{b/a}. \tag{3.13}$$

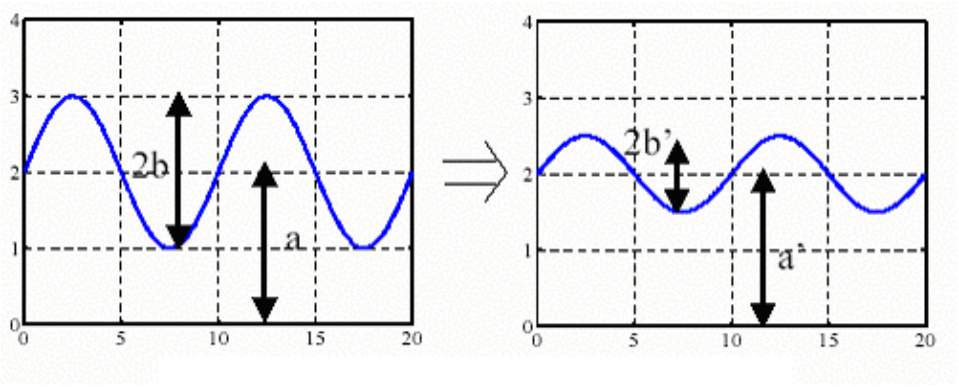


Fig. 22. (a) Input signal: $\Phi(x, f_i) = a + b\sin(2\pi * f_i x)$; (b)Output signal: $\Phi'(x, f_i) = a' + b'\sin(2\pi * f_i x)$.

MTF is dominated by the aberrated Optics MTF, which may due to the following factors [39]:

- Optics Degradations MTF.
- Jitter MTF.
- Smear MTF.

The total MTF is the product of all applicable MTFs given by

$$MTF_{total} = MTF_{diff} * MTF_{smear} * MTF_{alignment} * MTF_{jitter} * ...$$

The actual MTF may be much lower than the predicted analytically. Mostly, MTFs are obtained by photographing a chart (typically the USAF 1951 lens test chart) and looking for the highest resolution pattern where detail was visible. More optics knowledge can be found in [40, 41, 42]. It's very difficult to get an accurate mathematical frequency response equation for any given lens. We use the following expression to analytically describe the lens quality or the frequency response of the filter.

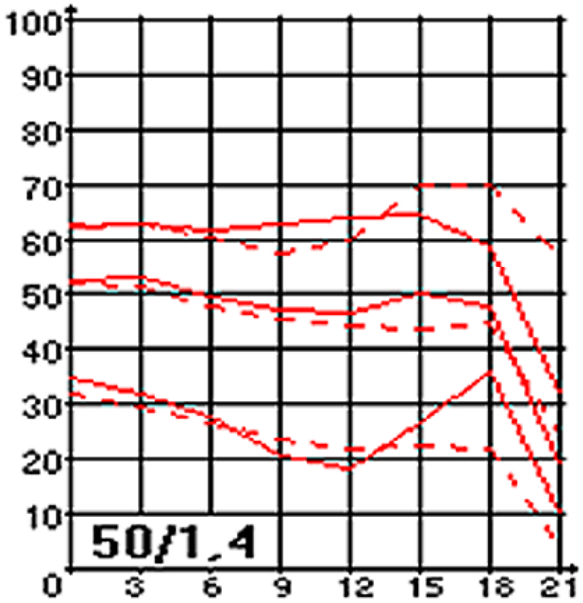


Fig. 23. Testing result of Pentax SMC-F 50/1.4 [2]. The graphs show MTF in percent for the three line frequencies of $10lp/mm$, $20lp/mm$ and $40lp/mm$, from the center of the image (shown at left) all the way to the corner (shown at right). The top two lines represent $10lp/mm$, the middle two lines $20lp/mm$ and the bottom two lines $40lp/mm$. The solid lines represent sagital MTF (lp/mm aligned like the spokes in a wheel). The broken lines represent tangential MTF (lp/mm arranged like the rim of a wheel, at right angles to sagital lines). On the scale at the bottom 0 represents the center of the image (on axis), and 21 represents 21 mm from the center, or the very corner of a 35 mm-film image.

$$h(x,y) = \frac{1}{2\pi\sigma_o^2} e^{(x^2+y^2)/\sigma_o^2}, \tag{3.14}$$

where the constant term σ_o account for the intrinsic blur due to finite lens resolution. The value of σ_o totally depends on the lens quality. The smaller the value of σ_o, the better the lens quality is. For an ideal thin lens, $\sigma_o = 0$, the cutoff frequency is infinitely high. Given in Fig. 23 [2] is the MTF graph of a typical commercial lens (Pentax SMC-F 50/1.4) that is used in our first testing prototype. From Fig. 23, we can see that this lens only keeps 50% contrast with $20lp/mm$ input resolution.

b. Sampling Imaging System

The ability of the camera system to accurately capture all of image details is also dependent upon the sampling interval. Features seen in the sensor that are smaller than the digital sampling interval (have a high spatial frequency) will not be represented accurately. The Nyquist criterion requires a sampling interval equal to twice the highest specimen spatial frequency to accurately preserve the spatial resolution in the resulting digital image. To ensure adequate sampling for high-resolution imaging, an interval of 2.5 to 3 samples for the smallest resolvable feature is suggested.

Since the photoreceptor has finite size, it is a spatially low-pass filter, whose cutoff frequency depends on the area of photoreceptor. According to the sampling theorem, the effect of the square-box spatial filter is given by a sinc function $\left(sinc(x) = \frac{sin(\pi x)}{\pi x}\right)$.

Based on the MTF graph given by Fig. 23, the lens attenuates the contrast of $20lp/mm$ signals by 50%. Considering the $3dB$ attenuation criterion for the low pass filter, we may regard the lens as an ideal low-pass filter with cutoff frequency $f_c = 20lp/mm$ to simplify the model. According to the sampling theorem, if the size of the photopixel is smaller than $25\mu m$, all the signals collected by the lens are preserved. Of course, this assumption is subjective and is selected based on how pleasing the imagery is to the focus measure processor. To minimize the error due to the assumption, the size of the photopixel can be decreased. For example, Schade criterion is a well-known criteria for the design of the sampled imaging system. Schade suggests that the pre-sample MTFs should be 0.4 or less at the half-sample rate to minimize the sampling artifacts [43]. Based on his criterion and Fig. 23, the sampling pixel size need to be smaller than $13\mu m$.

c. Focus Distribution and Detection

Referring to Fig. 24, and assuming that a planar object satisfying the occlusion condition is located at $z = z_o$ with a radiance distribution $R(\xi, \eta)$ on the local object coordinates ξ and η, the irradiance distribution $I(p, s)$ on the tilted sensor plane can be expressed generally as

$$I(p,s) = \frac{\pi D^2}{4z_o^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(p, s, p_\xi, s_\eta) R(\xi, \eta) d\xi d\eta, \tag{3.15}$$

where D is lens diameter, $h(p, s, p_\xi, s_\eta)$ is the point-spread function (PSF), and p_ξ, s_η denote the sensor-plane coordinates of the principle ray projected by an arbitrary object point $O(\xi, \eta)$. The simplest yet meaningful object satisfying the invariant-pattern condition is an edge oriented along η. Assuming a negative step radiance-function of magnitude R, i.e., $R(\xi, \eta) = Ru(-\xi)$, and transforming the integration variables ξ and η into $\imath_\xi$ and s_η through usual geometric relations, (3.15) can be converted into

$$I(p,s) = \frac{\pi R \sin\alpha}{4N^2} \int_{s_\eta=-\infty}^{\infty} \int_{p_\xi=0}^{\infty} h(p, s, p_\xi, s_\eta) R(\xi, \eta) d\xi d\eta, \tag{3.16}$$

where $N \equiv f/D$ is f-stop.

Adopting a symmetrical Gaussian PSF is very common in modeling defocus in a usual frontal camera [23, 20]. The center of the Gaussian is assumed to coincide with the point where the principle ray intersects the sensor plane, and the standard deviation is assumed to be proportional to the radius of blur circle. Extending this approach to our case, in which blur is elliptical (Appendix A), the PSF is described by

$$h(p, s, p_\xi, s_\eta) R(\xi, \eta) = \frac{1}{2\pi\sigma_p\sigma_s} exp\left[-\frac{1}{2}\left(\frac{p - p_\xi}{\sigma_p}\right)^2 - \frac{1}{2}\left(\frac{s - s_\eta}{\sigma_s}\right)^2\right], \tag{3.17}$$

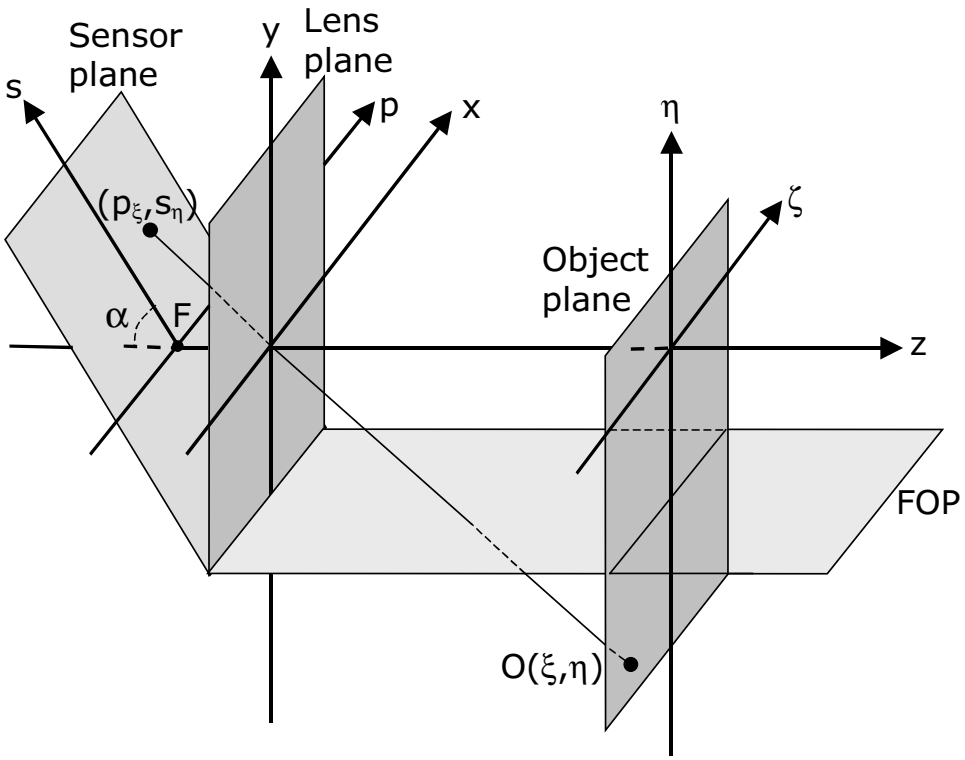


Fig. 24. The reference convention for the analytical model of image projection onto a tilted sensor plane. Note that introduction of object-plane coordinates ξ and η, whose origin is at $x = 0$, $y = 0$, $z = z_o$.

where σ_p and σ_s denote the standard deviations, which are proportional, respectively to the minor radius and major radius of the blur ellipse with a common constant of proportionality k. The radii, in turn, can be easily formulated with ray tracing, and shown to be proportional to $|s_\eta - s_i|$. This implies $\sigma_p \propto |s_\eta - s_i|$ and $\sigma_s \propto |s_\eta - s_i|$.

When point $O(0, \eta)$ moves down from FOP, its image in the tilted plane blurs. Shown in Fig. 25 is the sequence of the blurring ellipses as point O moves away from FOP along η-axis. Fig. 25 is generated by Matlab when camera setting are $f = 50mm$, $\alpha = 45^o$, f-stop=1.4, and $\sigma_o = 1\mu m$. The position of the focused row is $s = 1420\mu m$. The shaded pixels in Fig. 25 are photoreceptors, whose active area is $20\mu m^2$. Note that there is no gap between neighboring photoreceptors in the real chip. The row pitch is $120\mu m$. As we can see, spatial cutoff frequencies of the rows depend on their position along s-axis.

If, however, k is sufficiently small, then, the variation of h with s_η peaks very sharply around $s_\eta = s$, which enables us to assume σ_p and σ_s to be independent of s_η within this narrow neighborhood of s, and be represented with the values they attain for $s_\eta = s$. This makes an analytical integration of (3.16) possible, and yields

$$I(p, s) = \frac{\pi}{2} \frac{R \sin \alpha}{(2N)^2} [1 + erf(\frac{p}{\sqrt{2}\sigma_p})], \tag{3.18}$$

where, σ_p is given by

$$\sigma_p = \frac{k}{2N} \frac{\sin \alpha}{\sqrt{\tan^2 \alpha - (1/2N)^2}} |s - s_i| + \sigma_o. \tag{3.19}$$

Note that we assume $\tan \alpha > (1/2N)$ in deriving (3.18) from (3.16), and the intrinsic blur due to finite lens resolution σ_o is also included in (3.19).

We have put to test the analytical model described by (3.18)-(3.19) by experimentally analyzing the image of an edge as a function of N and α, which appear to be

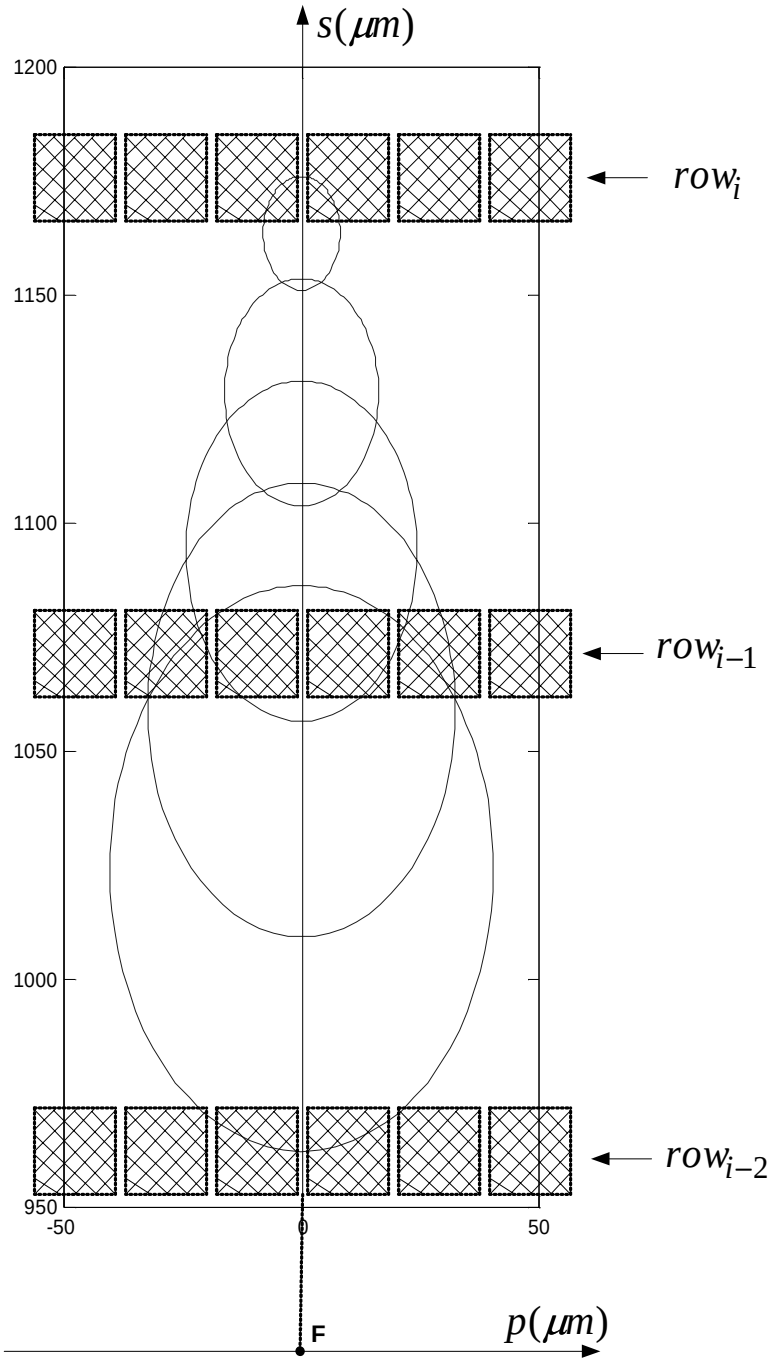


Fig. 25. The sequence of the blurring ellipses as point O moves down from FOP along η-axis. Fig. 25 is generated by Matlab when camera setting are $f = 50mm$, $\alpha = 45^o$, f-stop=1.4, and $\sigma_o = 1\mu m$. Note that the coordinates s and p are based on Fig. 24, σ_o is set to a very small value so that lens filtering effects are minimized.

the only adjustable camera parameters affecting the irradiance distribution. To verify that, two camera were built and used for experiments. Shown in Fig. 26 are three experimental images and the corresponding irradiance profiles along p. The profiles belong to the row of the best focus ($s = s_i$) and two additional rows equidistant from the latter. The validity of the qualitative features of the model is obvious from the plots which exhibit not only the predicted error-function distribution of irradiance with one-half of the maximum at the column of best focus, but also the expected symmetry with respect to $s = s_i$. A quantitative verification is possible by extracting k and σ_o, and inspecting their invariance to camera settings. We do this by measuring (a) the irradiance step ΔI, (b) the irradiance gradient $\frac{\Delta I}{\Delta p}$ at $p = 0$ once for $s = s_i$ and then for $s \ll s_i$, and using these values in the following expressions obtained from (3.18and (3.19).

$$\Delta I \equiv I(\infty, s) - I(-\infty, s) = \frac{\pi R \sin\alpha}{(2N)^2}, \tag{3.20}$$

$$\frac{\Delta I}{\Delta p} \equiv \begin{cases} \frac{\partial I(p,s)}{\partial p}|_{p=0, s \ll s_i} = \sqrt{\frac{\pi}{2}} \frac{R}{2N} \frac{\sqrt{\tan^2\alpha - (1/2N)^2}}{k|s - s_i|} \\ \frac{\partial I(p,s)}{\partial p}|_{p=0, s=s_i} = \sqrt{\frac{\pi}{2}} \frac{R \sin\alpha}{\sigma_o (2N)^2} \end{cases} \tag{3.21}$$

k and σ_o are calculated from (3.20) and (3.21) after canceling out R whose exact value is unknown due to uncalibrated illumination of the edge and the automatic gain control of the video chip. The values thus extracted from the entire set of experimental data are $k = 0.445 \pm 0.024$ and $\sigma_o = 12.1 \pm 2.1 \mu m$. The very tight distribution of these two parameters is indicative to the quantitative agreement between the model and experimental results.

d. Focus Measure Algorithm

Probably the most widely recognized focus measures are those based on image derivatives, L_1 or L_2 norm of gradient or second derivative are the most common, and

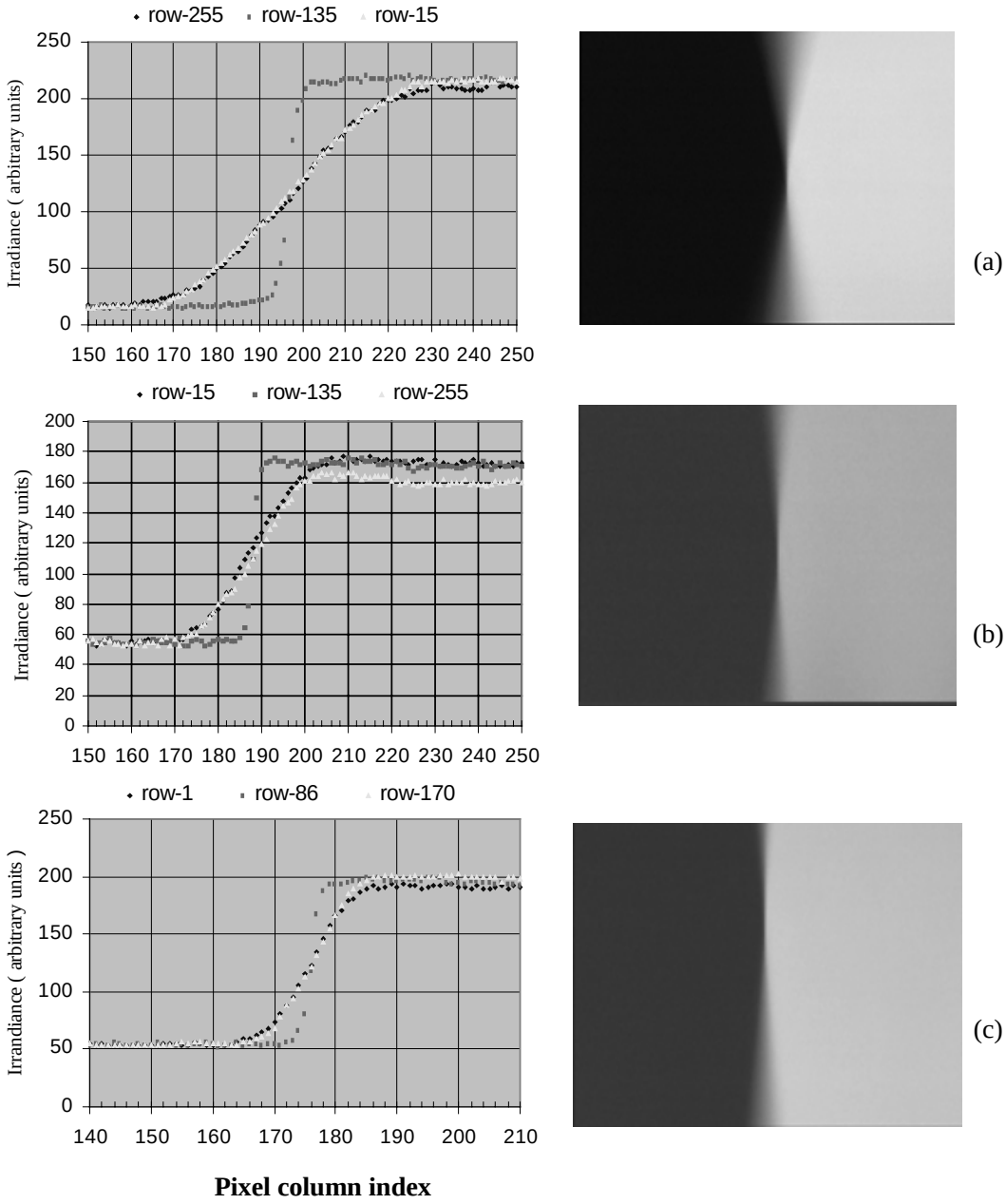


Fig. 26. Experimental images of an object edge located at $z_o = 207cm$. Also shown are the irradiance profiles along the row of best focus and along two farthest rows equidistant to the latter. Note that the irradiance step ΔI and the extent Δp of its spread along p are used for calculating the coefficient k and the constant σ_o. (a) $\alpha = 30^o$, $N = 1.4$. (b) $\alpha = 30^o$, $N = 2.8$, (c) $\alpha = 45^o$, $N = 2.8$.

are usually applied in two dimensional form. In our case, where real-time memoryless focus-detection by an embedded analog network is the goal, we opt for one-dimensional derivatives along p. This decision is influenced by our concern for hardware simplicity, and is highly justified by the pattern-invariance condition which calls for an object pattern comprised mostly edges oriented along y. Furthermore, we have decided to deploy logarithmically compressing photosensors in our custom sensor/processor chip in order to accommodate the wide dynamic range of natural irradiance. In our case, therefore, the signal roll-off with s around the point of best focus can be defined as

$$r_1 \equiv \left| \frac{1}{\left(\frac{\partial \ln I}{\partial p}\right)_{max}} \frac{d}{ds} \left(\frac{\partial \ln I}{\partial p}\right)_{max} \right|_{s=s_i} \tag{3.22}$$

for the gradient-based measure, and as

$$r_2 \equiv \left| \frac{1}{\left(\frac{\partial^2 \ln I}{\partial p^2}\right)_{max}} \frac{d}{ds} \left(\frac{\partial^2 \ln I}{\partial p^2}\right)_{max} \right|_{s=s_i} \tag{3.23}$$

for the second-derivative-based measure. According to 3.18-3.19, the maximum of the derivatives with respect to p occur approximately at $p = 0$ for (3.22) and at $p = \pm\sigma_p$ for (3.23), and the corresponding signal roll-off rates are approximately described by

$$r_1 = \frac{1}{2} \frac{k \sin\alpha}{N\sigma_o \sqrt{\tan^2\alpha - (1/2N)^2}}, \tag{3.24}$$

$$r_2 = \frac{k \sin\alpha}{N\sigma_o \sqrt{\tan^2\alpha - (1/2N)^2}}, \tag{3.25}$$

which clearly show the advantage of using a small f-stop, a small tilted angle and high-resolution optics. The factor of 2 improvement indicated by (3.25) shows the second-derivative-based focus measure is more suitable to pick up the focus winner. These equations yield $r_1 = 1.44$ percent per micrometer and $r_2 = 2.89$ percent per micrometer for the extracted values $k = 0.445, \sigma_o = 12.1\mu m$ and camera parameter

$N = 1.4$ and $\alpha = 30^o$. These numbers represent probably the best case of roll-off because there is little room for further improvement by decreasing N and σ_o without raising the cost of optics, or by tilting the angle more without lowering the irradiance to undetectable levels. In any case, the actual signal difference between the row of best focus and its neighbors may be significantly lesser than what these number indicate because the spatial integration of irradiance over photosensor aperture is not taken into account in deriving (3.24) and (3.25).

For a more common image captured by the discrete pixels, we compute the partial derivatives of pixel j in row i by [21]

$$ML(i,j) = \left| \, 2\ln I_{i,j} - \ln I_{i,(j-1)} - \ln I_{i,(j+1)} \right|. \tag{3.26}$$

Finally, the focus measure of row i is computed as is the sum of modified Laplacian (SML):

$$F(i) = \sum_{j=1}^{N} ML(i,j), \tag{3.27}$$

where N is the number of pixels in row i.

CHAPTER IV

IMAGE SENSING CIRCUIT

The proposed system is very suitable to be implemented with an analog sensor/processor chip. Monolithic solutions, offer advantages of the absence of leakage currents and other signal degradation caused by off-chip connection [44]. Because image irradiance is a continuous function of time, asynchronous circuit implementation is preferable to clocked implementation. Analog processor network is extremely suitable for this kind of signal processing. Its main drawback is its lack of precision due to the noisy input data. In this chapter, we discuss the low-level issues to do with photoreceptors.

Because our system introduces a tilted sensor plane, not all the light will be absorbed by the silicon and generate photocurrent. We start by discussing the light reflection on the interfaces of different materials. Then we discuss physical principles of the photoreceptors, and then view the specs of the traditional CCD and CMOS photoreceptors. In the end of this chapter, we propose and characterize the photoreceptor architecture used in our system.

A. Reflection, Refraction and Transmission

Almost all image sensors depend on the generation of electron-hole pairs when photons strike a suitable material. This is the basic process in biological vision as well as photography. However, not all incident photons generate an electron-hole pair. Some pass right through the sensing layer, some are reflected, and others lost energy in different ways. Since the sensor used in our system is tilted, part of the light is reflected back to the air. Only part of photons can enter the silicon layer and create electron-hole pair. This section discussed reflection phenomenon on the surface of silicon pixels.

In general, when an incident-plane wave arrives at the interface between two media of refractive indices n_i and n_t, part of the incoming wave is reflected back into the incident media (the reflected wave), while the remainder propagates into the transmitting media (the refracted or transmitted wave).

Shown in Fig. 27 is a ray diagram depicting such an interaction, where the subscripts "i", "r", "t" refer to incident, reflected and transmitted respectively. The frequency of the reflected and transmitted waves are the same as that of the incident wave, but the speed of the transmitted wave differs from that of the incident wave and therefore has a different wavelength.

The three basic laws of reflection and refraction are as follows:

- The incident, reflected and transmitted rays all reside in a plane, known as the plane of incidence, which is normal to the surface.
- The angle of incident is equal to the angle of reflection: $\theta_i = \theta_t$.
- The incident and transmitted ray directions are related by Snell's law: $n_i \sin\theta_i = n_t \sin\theta_t$.

The fraction of energy reflected at the interface of the two media depends on the indices of refraction of the two media, as well as on the angle of incidence and the polarization of the incident wave. When electromagnetic radiation in the light region is incident normally, the reflected intensity Ir can be found by

$$I_r = I_0 \frac{(n_t - n_i)^2}{(n_t + n_i)^2} \tag{4.1}$$

Where I_0 is the incident intensity and n_i and n_t are the indices of refraction of the two media. Intuitively we can see that the large θ_i, the more photons reflected back to air without being absorbed by silicon.

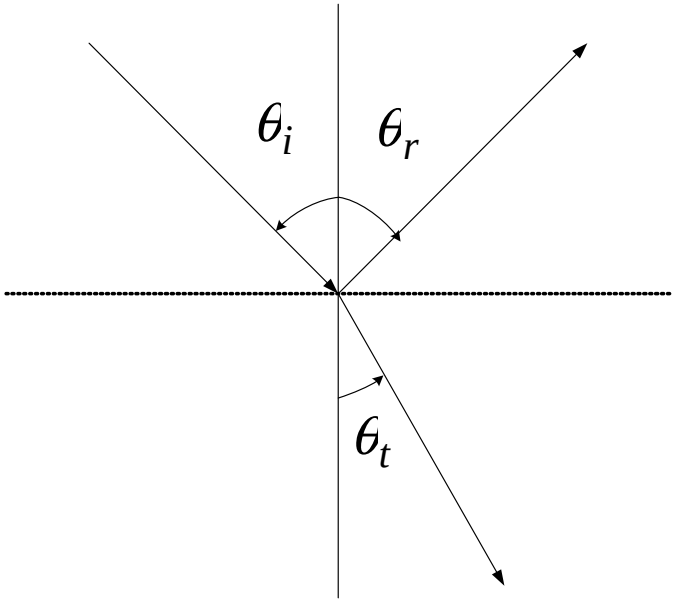


Fig. 27. Reflection and refraction of an incident plane wave arriving at the interface between two media of refractive indices n_i and n_t.

Absorption

Absorption of photons refers to their attenuation by the process of conversion to other forms of energy. When a photon strikes an atom, part or all of its energy is transferred to that atom, depending on the energy of the initial photon. Only the photons absorbed by the silicon can create electron-hole pairs.

B. Principle of Photodetectors

Photodetectors are semiconductor devices that can convert optical signals into electrical signals. The operation of a photodetector involves three steps: (1) carrier generation by incident light, (2) carrier transport and/or multiplication by whatever current gain mechanism, (3) interaction of current with the external circuit to provide the output signal.

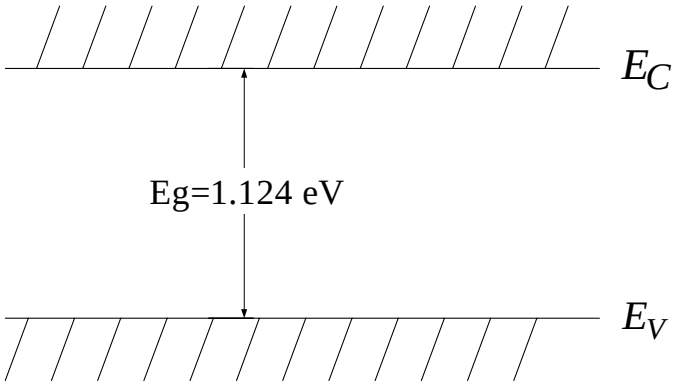


Fig. 28. Diagram of silicon energy bands.

1. Radiative Transitions

There are three processes for interaction between a photon and an electron: absorption, spontaneous emission, and stimulated emission.

Consider two energy levels E_v and E_c of an atom in Fig. 28, where E_v is in the valence band, E_c is in the conduct band. When an electron in state $E1$ absorbs a photon of energy equal to $(E_c - E_v)$, it goes to E_c. The change of the energy state is the absorption process. Si ($E_g = 1.12eV$) is the semiconductor material usually used to fabricate photo-conductors intended for visible-light. The maximum useful photon wavelength for a semiconductor of band gap E_g is given by $\lambda_{max} = hc/E_g$, where $h = 4.135 \times 10^{-15} eV$ is Planck's constant, $c = 3 \times 10^8 m/s$ is the speed of light, and λ is the wavelength of light.

Coincidently, it's luckily that the photon energy of visible light is large enough to create electron-hole pair. The following list is wavelength and photon energy of visible light [32].

- $\lambda = 400nm$ corresponds to violet $E_{ph} = 3.1eV$
- $\lambda = 480nm$ corresponds to blue $E_{ph} = 2.58eV$

- $\lambda = 540nm$ corresponds to green $E_{ph} = 2.3eV$
- $\lambda = 700nm$ corresponds to red $E_{ph} = 1.77eV$

When the energy of an incident-light photon E_{ph} exceeds the semiconductor's band-gap value E_g , absorption of the photon can take place through the excitation of a valence electron into the conduction band, in addition, the excess energy $(E_{ph} - E_g)$ is dissipated as heat. The absorption coefficient is defined as the relative rate of decrease in light intensity along its propagation path. The absorption coefficient for a light photon of energy hv is proportional to the probability of a transition of an electron from the initial state to the final state to the final state (P_{rif}), to the density of electrons in the initial state (n_{init}) and to the density of final state (n_{final}). So the absorption coefficient can be calculated as $\alpha = \Sigma P_{rif} n_{init} n_{final}$ [45].

The photon flux at a depth x from the surface of a photoconductor generated through irradiation by a photon source (with a photon flux ϕ_0 given by $\phi_0 = M\lambda/hc$) is given by $\phi(x) = \eta\phi_0 exp(-\alpha x)$. For total absorption of incident photons, the detector thickness d must be several times greater than the inverse of the absorption coefficient α^{-1}. In such cases, the steady-state generation of carriers per unit area is given by $g = \eta\phi_0[1 - exp(-\alpha d)]$, where η is the internal quantum efficiency.

Photons with different wavelengths will be absorbed with different depths [32]:

- 99% of violet and blue light absorbed with depth of $0.6\mu m$.
- Need $x \approx 16.6\mu m$ for the same red light absorption.
- These depths are quite consistent with the junction depth of a CMOS process.

The wavelength-dependent absorption means that photodetectors formed from junctions with different junction depths will have different spectral responses. In a standard n-well CMOS process, there are three types of junctions: the n-well and

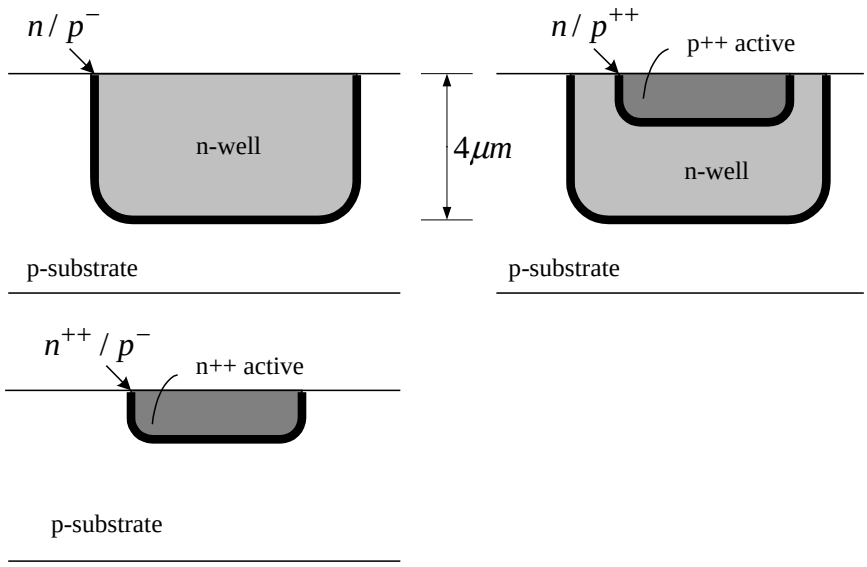


Fig. 29. The junctions in standard n-well CMOS process.

p-substrate junction; the n-active and p-substrate junction; the p-active and n-well junction. Fig. 29 shows all the three junctions in standard n-well CMOS process. Note that the depth of well is about $4\mu m$ in model CMOS process.

Silicon, the major material used in photovoltaic cells, is an indirect bandgap material. This means that in order to make the transition between the valence and conduction band, we must change the kinetic energy of the electron as well as the potential.

The quantum efficiency defines the number of excess carriers generated per incident photon. The reflection at the surface of the air-semiconductor interface is due to the refractive index difference, as well the energy of the incident radiation and the sensitive volume of the detector.

The quantum efficiency, defined as the number of excess carriers generated per incident photon, is in the case of photodiodes given by

$$\eta(\lambda) = \frac{\#e - h\ pairs}{\#incident\ photons\ at\ wavelength\ \lambda} \tag{4.2}$$

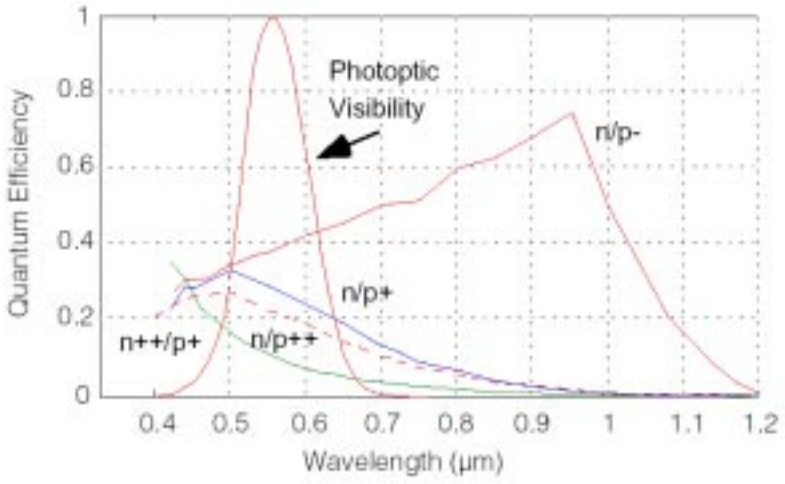


Fig. 30. Measured spectral quantum efficiency versus photon wavelength. Each curve is labeled with the name of the junction [3].

where #e-h pair is the electron-hole pair generated by a source of photons of energy with the wavelength λ. One of the key factors that determine η is the absorption coefficient α. Since α is a strong function of the wavelength, the wavelength range in which appreciable photocurrent can be generated is limited. The long-wavelength cutoff λ_c is established by the bandgap. The short wavelength cutoff of the photo response comes about because for short wavelengths the values of a are very large, and hence the radiation is mostly absorbed very near the surface where recombination time is short. Therefore, the carriers can recombine before they can be collected in the p-n junction. From now on, we mainly use quantum efficiency to analyze the photodetectors instead of using the absorption coefficient. The dependence of quantum efficiency on photon wavelength is shown in Fig. 30.

Factors playing a role in the determination of the response speed include the drift time within the depletion region, the diffusion time to the depletion region of carriers generated outside the depletion region and the width and capacitance of the depletion region. Contributions to the device noise include shot noise (originating

from randomness in the generation of the reverse current and signal current), thermal noise (arising from the junction resistance) and quantum noise (associated with the optical signal itself).

2. Review of Silicon Photosensors

a. Historical Background

Before CMOS sensor and before CCD's, there were MOS image sensors. In the 1960's there were groups working on solid-state image sensors using NMOS, PMOS, and bipolar processes. In 1963, Morrison reported a structure that allowed determination of alight spot's position using the photoconductivity effect. In 1966, Westinghouse reported a 50×50 element monolithic array of phototransistors [46]. All of these sensors had an output signal proportional to the instantaneous local incident light intensity and did not perform any intentional integration of the optical signal. Therefore, the sensitivity of the sensors was low and they required gain within pixel to enhance their performance.

In 1967, Weckler at Fairchild suggested operating p-n junctions in a photon flux-integrating mode. The photocurrent from the junction is integrated on a reverse-biased p-n junction capacitance. Readout of the integrated charge using a PMOS switch was suggested. The issue of fixed-pattern noise (FPN) was explored in 1970 by Fry, Noble and Rycroft.

Until recently, FPN has been considered as the primary problem with MOS and CMOS image sensors. In 1970, when CCD was first reported, its relative freedom from FPN was one of the major reasons of its adoption over many other solid-state image sensors.

In the early 1990's, two independent efforts have led to resurgence in CMOS

image sensor development. The first effort was to create highly functional single-chip imaging systems where low cost, not performance, was the driving factor. The second effort grew from the need for highly miniaturized, low-power, instrument imaging systems.

b. The CCD Approach

CCD technology was developed for imaging applications, and its fabrication processes were optimized to build an image sensor with the best possible optical properties and image quality. CCD technology is mature with respect to production yield and performance. Both benchmarks are at either theoretical limits or practical levels significantly unchanged for several years.

A CCD is composed of pixels, or picture elements, arranged in an X, Y matrix consisting of rows and columns. Each pixel, in turn, is composed of a photodiode and an adjacent charge transfer region, which is shielded from light. Adjacent charge transfer regions are arranged in a column to form a vertical charge transfer register. The photodiode converts light (photons) into charge (electrons). The number of electrons collected is proportional to the light intensity.

A CCD is a dynamic analog (charge) shifter register implemented using close spaced MOS capacitors. Typically, light is collected over the entire imager simultaneously and then transferred to the adjacent charge transfer cells in the columns. Next, the charge must be read out. To do this, one row of data, including signals from one pixel in each of the columns, is transferred from the vertical charge transfer register to a separate horizontal charge transfer register. The charge packets for a given row are then read out serially and sensed by a charge-to out and an image can be displayed.

This architecture produces a low-noise, high-performance imager, but it has tradeoffs in terms of the manufacturing process. For example, CCD process tech-

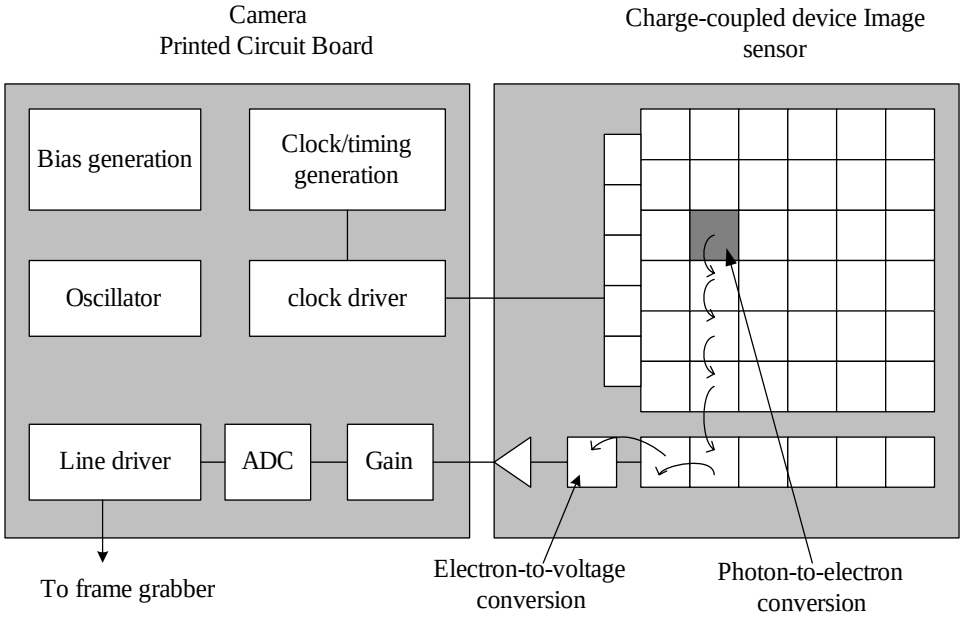


Fig. 31. Architecture of CCD camera.

nologies have been optimized to improve image quality, but in so doing, the processes are now unsuitable for efficient integration of other electronics onto the silicon.

Operating a CCD also requires application of several clock signals, clock levels, and bias voltages, complicating system integration and increasing power consumption, system bulk, and cost.-voltage conversion and amplifier section (see Fig. 31 [47]). The next row of data is then clocked into the horizontal transfer register. The process is repeated until all rows are read.

The CCD transfer noise is negligible because modern CCDs use buried channels. Without off-chip correlated double sampling, the predicted read noise is about $40e-$, nearly independent of array size. By applying CDS, the read noise lowers via kT/C noise reduction but becomes dependent on array size and video frequency.

CMOS imagers, on the other hand, are made with standard CMOS silicon processes in high-volume wafer fabs that produce ICs, such as microprocessors, microcontrollers, and DSPs. Therefore, the CMOS's pixel array can be formed on the

same device with standard electronics, such as digital logic, clock drivers, or A/D converters-a big advantage over the CCD processes. CMOS imagers can use the high-volume infrastructure of the semiconductor industry and will directly benefit from the progression of mainstream semiconductor technology, taking advantage of the move to smaller design rules and the ability to scale the technology to high volumes and resolutions.

To achieve these benefits, the CMOS imager architecture is arranged more like a memory cell or a flat-panel display. Each pixel contains a photodiode, which converts light to electrons; charge-to-voltage conversion section; reset and select transistor; and amplifier section.

Overlaying the entire pixel array is a grid of metal interconnects, which applies timing and readout signals, and an output signal metal interconnection for each column. The column output signal is connected to a set of decode and readout electronics, which are arranged for each column outside the pixel array. This architecture allows the pixel signals from the entire array, from subsections to individual pixels, to be read by a simple X, Y addressing technique -impossible with a CCD.

Table IV summarizes some of the main differences between the CCD and CMOS imager architectures.

The key difference between CMOS and CCD technology is the ability to integrate additional logic and achieve a camera on a chip. CMOS allows the consolidation of multiple discrete-logic and mixed-signal ICs in one device, reducing the size, part count, power consumption, and cost of the imaging solution. Many CMOS imagers don't perform at the same level as CCD imagers. Most notably, CMOS imagers can have high fixed-pattern noise, low sensitivity to light, high dark current, focal plane shutter effects, and some difficulty scaling to smaller pixel sizes.

There are several reasons to use CMOS pixel instead of CCD in our system.

Table IV. Comparison of CCD and CMOS image sensor features

CCD	CMOS
Smallest pixel size	Single power supply
Lowest noise	Single master clock
Lowest dark current	Low power consumption
100% fill factor for full-frame CCD	X,Y addressing and subsampling
Established technology market base	Smallest system size
Highest sensitivity	Easy integration of circuitry

- CCDs use a special manufacturing process to create the ability to transport charge across the chip without distortion. CMOS chips, on the other hand, use completely normal manufacturing processes to create the chip, which is cheaper.
- CCD doesn't have the ability to integrate sensors with analog and logic circuit on the same die.
- CCDs, use a special process that consumes lots of power. CCDs consume as much as 100 times more power than an equivalent CMOS sensor.
- The focus measure processor needs to access all photodetectors at the same time. CCD doesn't have this feasibility.

c. The CMOS Photodiode Approach

The increased demand for cheap consumer camera has led growing interests on CMOS image sensors [48]. Shown in Fig. 32 is the basic photodiode operation. The junction between n and p type silicon creates an electric field that separates photon-generated electron-hole pairs and counteracts the diffusion of carriers from the majority region

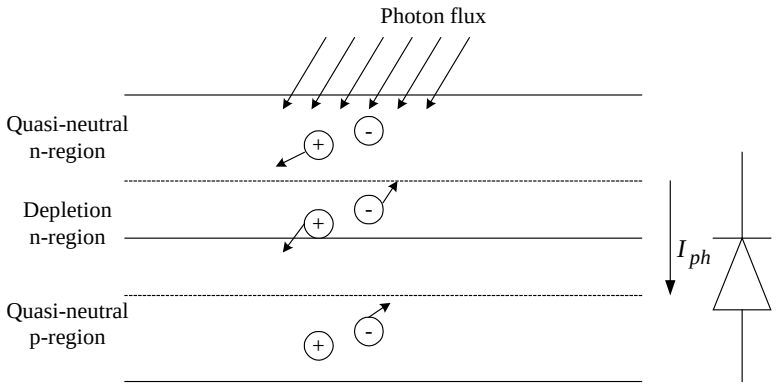


Fig. 32. The basic operation of photodiodes.

to the minority region. When a photon creates an electron-hole pair somewhere inside the junction, the electron and pairs are split apart by the junction field. When the pair is generated somewhere in the bulk region, the majority carrier is lost in the sea of majority carriers. The minority carrier starts diffusing. Two fates can occur: either the minority carrier recombines with a majority carrier, in which case it is as though the photon were never absorbed, or the minority diffuses to the junction and is swept across to the other side.

Photo current I_{ph} is the sum of three components:

- Current due to the generation in depletion region; almost all the carriers are swept by the strong field.
- Current due to holes in N-region; some diffuse to the depletion region and get collected.
- Current due to electrons in P-region.

In order to increase the photocurrent, we can try the following ways:

- Shallow the p-n junction so that more short-wavelength photons can enter the junction before being absorbed by the bulk.

- Increase the width of depletion region.

The depletion width W_{dep} is given by

$$W_{dep} = \sqrt{\left(\frac{2\varepsilon_s}{q}\right)\left(\frac{1}{N_A}+\frac{1}{N_D}\right)(V_0+V_R)}, \tag{4.3}$$

where $V_0 = V_T \ln \sqrt{\frac{N_A N_D}{{n_i}^2}}$ and V_R is the reverse bias voltage.

From Equation 4.3, we can see that there are two ways to increase the depletion width:

- Increase the reverse bias voltage V_R.
- Lighten the doping.

One drawback of increasing V_R is that it will increase the dark current of photodiode. Dark current is the leakage current of the photodiode, i.e. the current not generated by photo generation. It's called dark current since it corresponds to the photocurrent under no illumination. Dark current can't be determined analytically or using device simulation tools – can only be determined experimentally.

Absolute Current Level

A frequent asked question is the expected current with a given light intensity, a given photodiode area. For example, office fluorescent lighting conditions are an irradiance of $1W/m^2$, corresponding to an illuminance of 680 lux if the light is at the illuminance wavelength 555 nm. Under these conditions, a $20 \times 20\mu m^2$ photodiode with quantum efficient 0.5 generates a current equals to [3]

$$1\frac{J/s}{m^2} \times \frac{eV}{1.6\times 10^{-19}J} \times \frac{quantum}{2.5eV} \times (20\mu m)^2 \times 0.5 = 100pA \tag{4.4}$$

Typical sunlight is about 3 decades larger than the office lumination, corresponding to $100nA$ photocurrent.

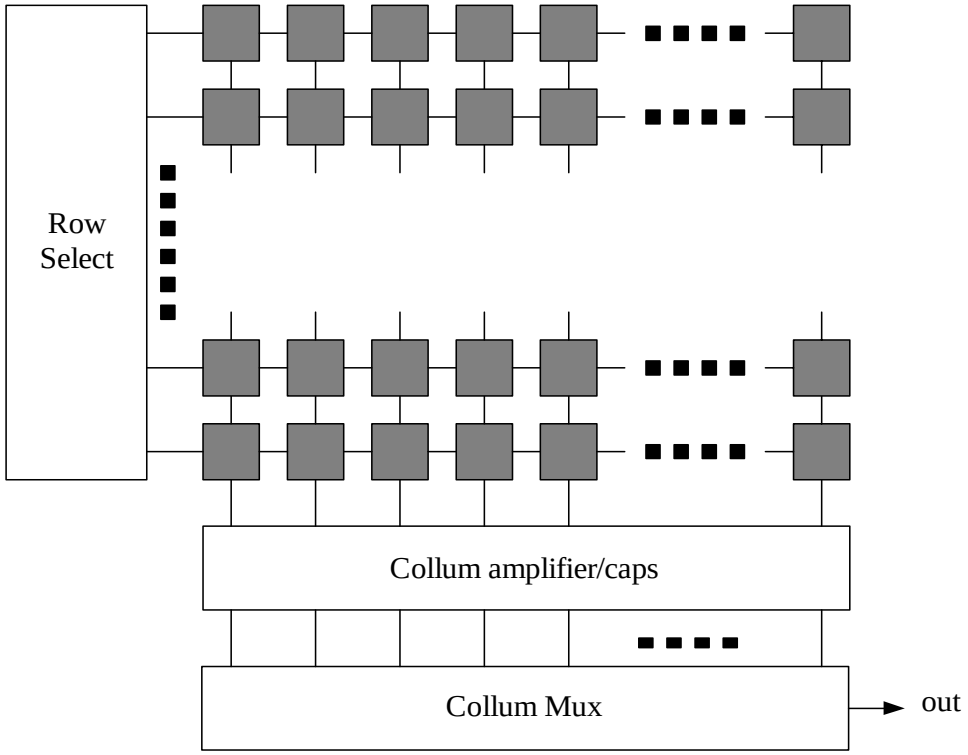


Fig. 33. The architecture of CMOS image sensor circuit.

d. Classical Passive and Active Photosensor

Given in Fig. 33 is the overall architecture of a classical CMOS sensor. The image sensor consists of an array of pixels that are typically selected a row at a time by row select logic. The pixels are read out to vertical column busses that connect the selected row of pixels to a bank of column amplifier and Mux. Row integration times are staggered by row/column readout time.

Photopixel circuits can be further divided into two categories: passive pixels and active pixels.

The passive pixel was first suggested by Weckler in 1967 [49]. Shown in Fig. 34 is the schematic and potential well of passive pixel. When the access transistor is activated, the photodiode is connected to a vertical column bus. A charge-integrating

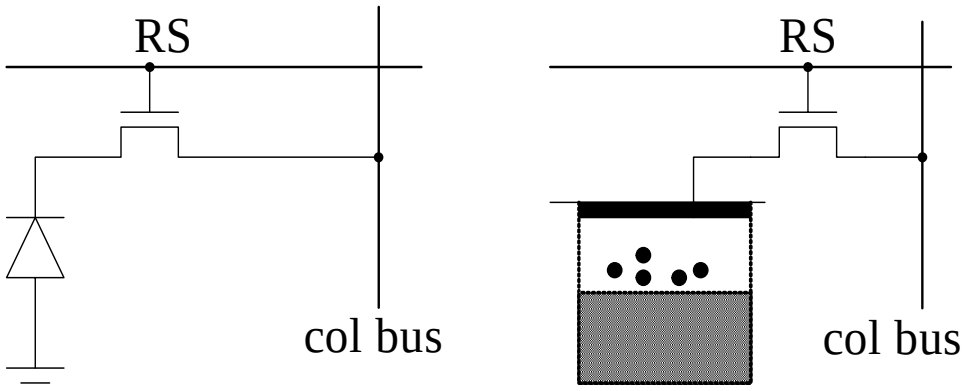


Fig. 34. Passive pixel schematic and potential well. When row select (RS) is pulsed, photo generated charge integrated on the photodiode is shared on the column bus.

amplifier read out circuit at the bottom of the column bus keeps the voltage on the column bus constant and reduces KTC noise [50]. When the photodiode is accessed, the voltage on the photodiode is reset to the column bus voltage. The advantage of passive pixels is the feasibility for achieving arrays with high quantum efficiency [51].

After the passive pixel was invented, it was recognized that the insertion of a buffer into the pixel could potentially improve the performance of the pixel. A sensor with an active amplifier within each pixel is referred to as an active pixel. A diagram of active photodiode is shown in Fig. 35. The first high-performance active photodiode sensor was demonstrated by JPL in 1995 in a 128×128 element array that had on-chip timing, control, correlated double sampling and fixed pattern noise suppression circuitry [52].

Now let's use passive pixel as an example to show the photo-to-voltage readout schemes in the integration mode. The output voltage of passive pixel sensor (PPS) is shown in Fig. 36. In steady state, assuming charge Q accumulated on the photodiode

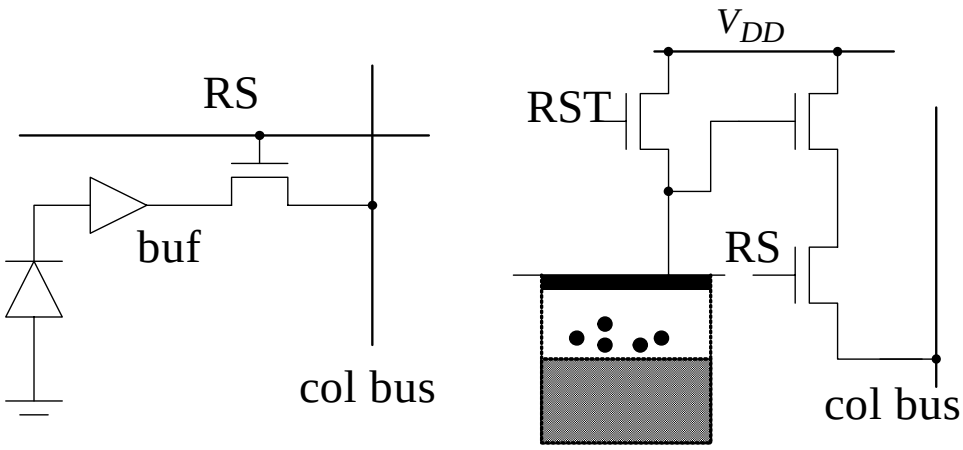


Fig. 35. Active pixel schematic and potential well. When row select (RS) is pulsed, the voltage on the photodiode is buffered by the source follower to the column bus. The photodiode is reset by transistor RST.

at the end of integration, the output voltage

$$V_o = V_ref + \frac{1}{C_f}Q \tag{4.5}$$

where $Q = i_{photo} \times t_{int}$, thus the sensor conversion gain is q/C_f (typically reported in $\mu V/electron$). From (4.5), with a fixed t_{int} and neglecting noise, V_o is proportional to i_{ph}, which is proportional to illumination.

Although the classical CMOS sensor structures are quite mature, they are not suitable for our system due to the following reasons:

(1)Limited dynamic range

Dynamic range quantifies the ability of a sensor to adequately image both high lights and dark shadows in a scene. It is defined as the ratio of the largest nonsaturating input signal to the smallest detectable input signal.

- The largest nonsaturating current is given by $i_{max} = qQ_{max}/t_{int} - i_{dc}$, where Q_{max} is the maximum Q in (4.5).

- The smallest detectable input signal is usually defined as the standard deviation

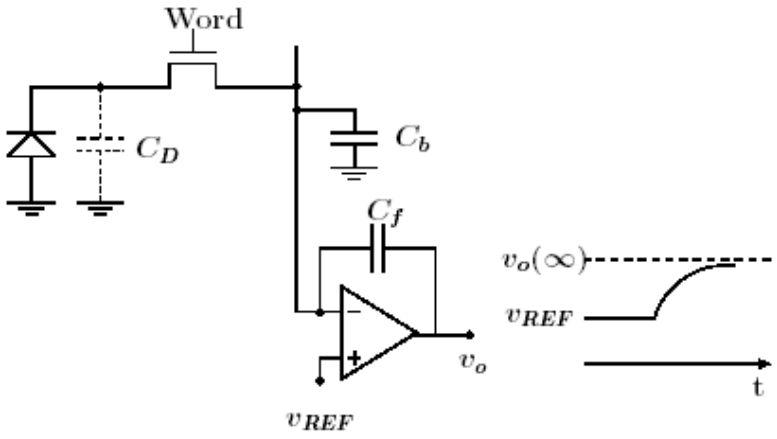


Fig. 36. PPS charge to voltage transfer function.

of the input referred noise under dark conditions In $\sigma_{In}(0)$ (the zero here refers to $i_{ph} = 0$), which gives $i_{min} = q/t_{int}\sqrt{i_{dc}t_{int}/q + {\sigma_r}^2}$, where ${\sigma_r}^2$ is the variance of read out noise in $electrons^2$.

Thus, the dynamic range can be represented in

$$DR = 20\log_{10}\frac{i_{max}}{i_{min}} = 20\log_{10} qQ_{max}/t_{int} - i_{dc}q/t_{int}\sqrt{i_{dc}t_{int}/q + {\sigma_r}^2} \tag{4.6}$$

Even if the adapted integration time scheme is used, which greatly enhanced the complexity of the circuit, the useful dynamic range is still poor. Table V lists some typical dynamic ranges. We can see that the dynamic range of the integration mode CMOS sensors is not large enough for our system that is designed for most natural scenes.

(2)Not a continuous time readout scheme

The operation and reading in each pixel is: reset, integration, reading and reset [53].

(3)Need ADC/DAC and memory

Since the pixel outputs are readout sequentially, the signal processing can't be

Table V. Dynamic range of image sensors

Image capture device or scenes	Dynamic range
Natural scenes	$> 100dB$
Human eye	Around $90dB$
Film	$80dB$
High end CCDs	$> 78dB$
Consumer grade CCD	$66dB$
Consumer grade CMOS sensors	54 dB

done locally. The image of the scene captures by the sensor has to be stored in memory before it can be processed.

e. The Continuous Time CMOS Photodiode

As we can see from above, the integration mode CMOS sensor is unsuitable for our system. What our system needs is a sensitive, continuous-time, logarithmic photosensor circuit. A logarithmic sensor is sensitive to relative changes in the intensity, not absolute intensity. Therefore, it is useful for reporting about image contrast. Image contrast is due mostly to the reflectance of the physical surfaces (aside from shadow). Logarithmic sensors are sensitive to properties of the surface, and not the lightning conditions – that is why they are useful in our system.

The passive photodiode is the simplest photodiode that convert photocurrent to voltage directly on the diode. For example, in the n++/p- junction in Fig. 29, when light shines on the silicon, it makes electron-hole pairs. When electrons freed in the p- substrate diffuse to the junction, they are swept home by the junction field into the n++ region. The result of the photocurrent flowing from n++ to p- is that the

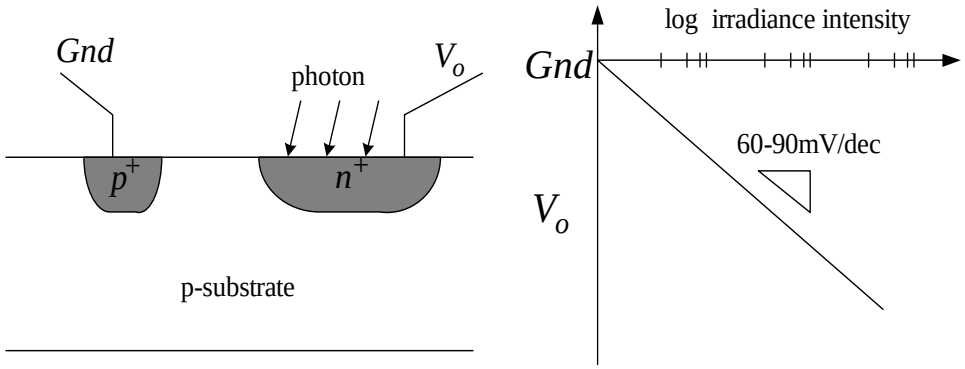


Fig. 37. Operation of a simple logarithmic passive photodiode.

n++ region becomes negatively charged with respect to the substrate. Since there is no current flowing in the diode, this negative voltage sets up a forward current in the junction to compensate for the photocurrent. In the forward region the iv relationship is closely approximated by [54]

$$i = I_s(e^{v/nV_T} - 1) \tag{4.7}$$

Since the forward current is exponential in the junction voltage, the voltage on the n++ region is logarithmic in the intensity. This relationship can be expressed alternatively in the logarithmic form:

$$v = nV_T \ln (I_{ph}/I_S) \tag{4.8}$$

This equation simply states for a decade (factor of 10) change in current, the diode voltage change by $2.3nV_T$, which is approximately 90 mV for $n = 1.5$. The typical value of I_S is $10^{14} - 10^{16}A$ [55].

However, as shown in Fig. 37, this signal is below the substrate voltage (Gnd), which means it's impossible to do any signal processing using such a low voltage. There are two ways to solve this problem. The first way is to use the p+/n-well junction. Shown in Fig. 38(a) is a passive way: the voltage of n-well V_b is set to a

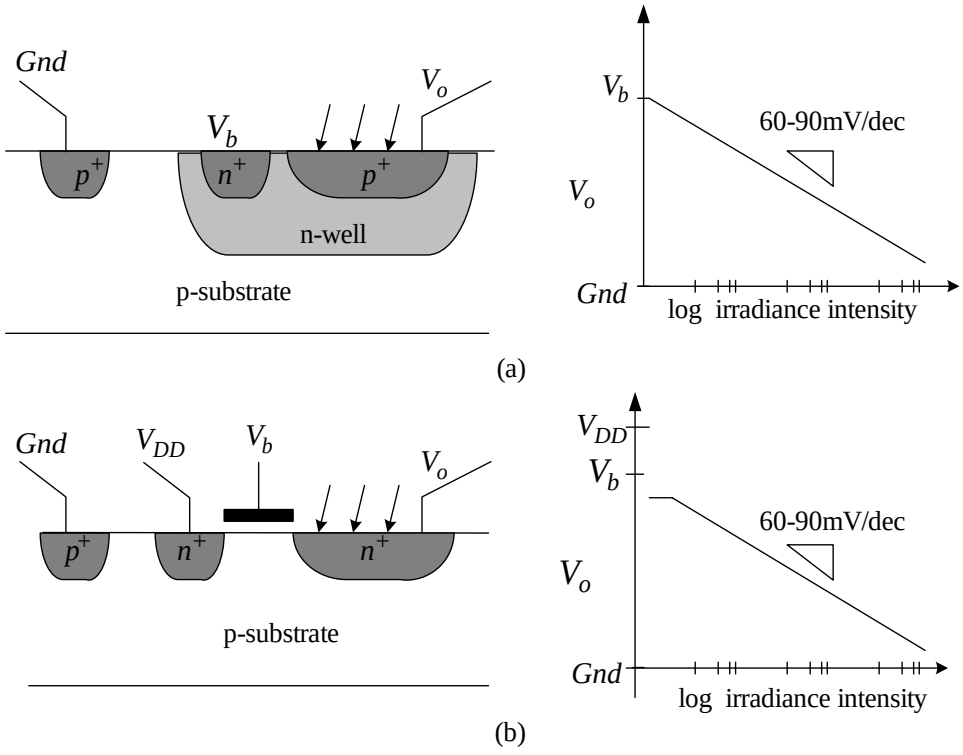


Fig. 38. The logarithmic photodiodes that can generate useful signals. (a) p++/n-well diode. (b) Diode driven by an active transistor working in subthreshold region.

voltage higher than Gnd. Shown in Fig. 38(b) is an active way: the gate voltage V_b is set to a voltage higher than Gnd. In both cases, the photodiode can generate a useful output voltage. The gain of about $60-90mV/decade$ results in a typical range of output ranges of perhaps $20-30mV$ from a natural scene. The architecture in Fig. 38(b) is commonly used in industry.

Logarithmic photodiodes have two common problems: (1) Mismatches. The differences between supposedly identical receptor outputs might be quite large comparing with the typical signal variations produced by real scenes. (2) Diffusion of minority carriers. A parasitic current is caused by optically generated electrons in the substrate which can diffuse to the neighboring pixels as shown Fig. 39. This parasitic current has damaging effects in our system because (a) The sensor used in our

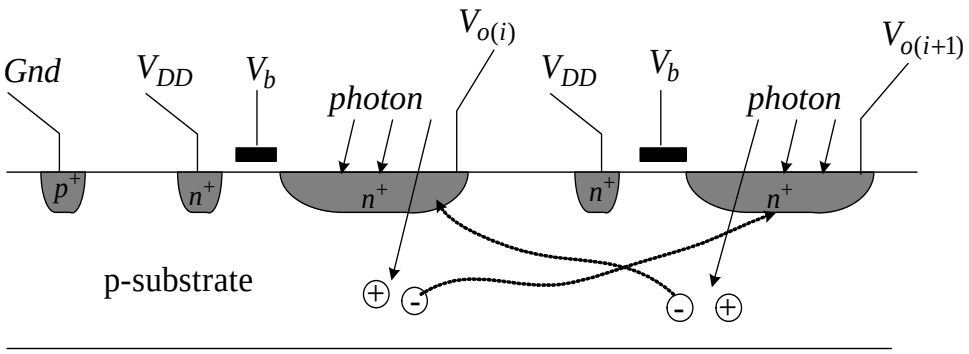


Fig. 39. The diffusion of optically generated electrons to neighboring pixels.

system is tilted regarding to the optical axis, which means that it may be easier for the electrons generated in the substrate under a pixel to diffuse to its neighbors. (b) The focus measure in our system relies on the output difference between the neighboring pixels.

In his PhD dissertation, *Delbrück* carefully analyze this problem. The results of his measurements show that the diffusing current is a function of the distance of the test spot from the center of the sensing pixel. For distances greater than approximately $70\mu m$, the decay of carrier concentration is exponential and the measured e-fold distance is about $30\mu m$. With no guard bar, the measured current is reduced by a factor of 10 in a distance of $40\mu m$. This result told us that to minimize the parasitic current, the distance of neighboring pixels should be larger than $40\mu m$. *Delbrück* uses $17\mu m$ wide wells to reduce the minority density by up to a factor of 10. Considering the effects of the tilted sensor, we need wider guard bars, which in turn waste the silicon area and reduce the sampling frequency.

To solve the above two problems, we use the architecture shown in Fig. 40.

It is built with a $p^+/n-well$ photodiode loaded with a series of three p-channel MOSFETs. Assuming MOSFET subthreshold operation, the output voltage V_o is

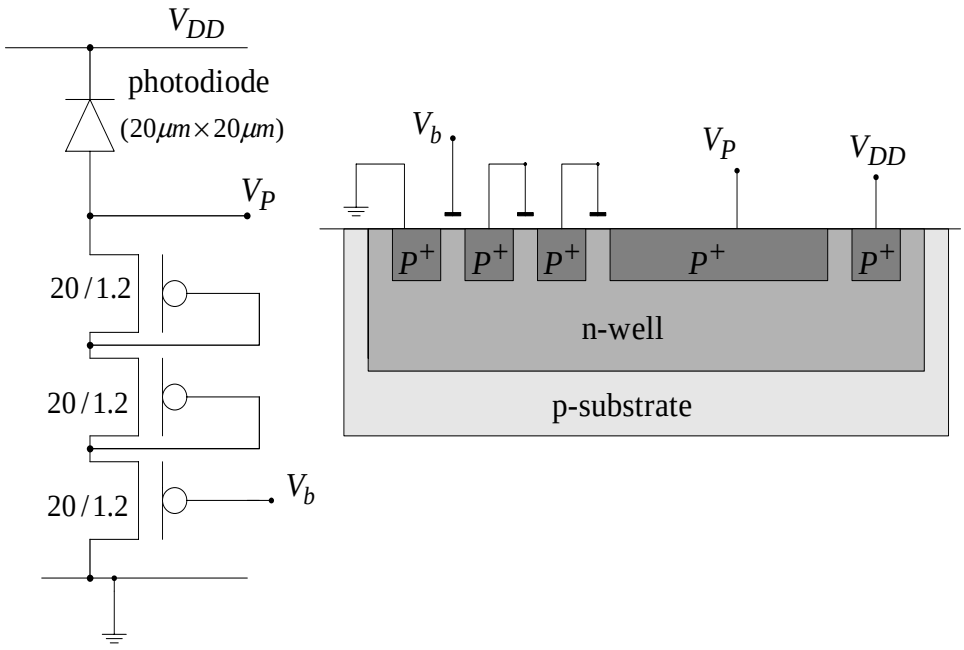


Fig. 40. The photodiode circuit used in our system.

described by

$$V_P = V_b + \frac{3s}{\ln 10} \ln(\frac{i_P}{I_0}) \tag{4.9}$$

where V_b is a level-shifting bias voltage, i_P the irradiance-proportional photocurrent, I_0 the MOSFET zero-bias current, and s the MOSFET subthreshold slope. Considering $s \approx 60mV/decade$, 4.9 indicates a logarithmic conversion of i_P, hence irradiance, into V_P at a rate around $180mV/decade$. The performance of the photosensor is shown in Fig. 41 and Fig. 42. The reason to introduce the level-shifting bias voltage V_b is based on the following observation: Although the environmental luminance can vary up to $5-6$ decades, the contrast of typical natural scene is less than 2 decades [3], which means the common-mode photovoltage of neighboring pixels may vary up to $1.2V$. This level-shifting bias voltage can adjust the range of V_P so that they won't saturate the inputs of analog focus processor.

This architecture uses the p^+/n-well diode instead of the traditional n^+/p^- sub-

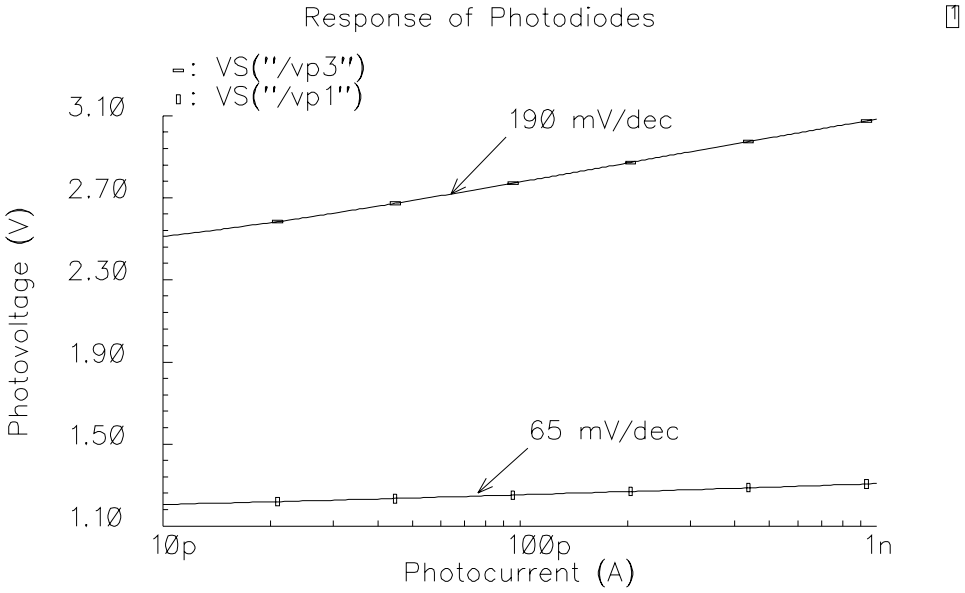


Fig. 41. Simulation result of the photopixel circuit used in our system compared with a traditional logarithmic photosensor. "vp3" is the output of our photodiode; "vp1" is the output of the traditional photosensor. Output voltage increases logarithmically with light intensity.

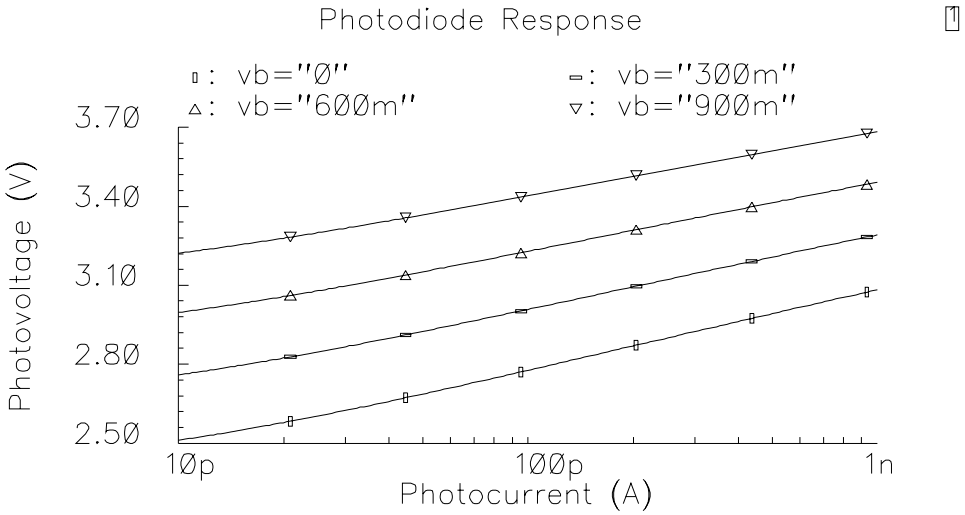


Fig. 42. Simulation result of the photopixel circuit used in our system. V_b shifts the output voltage.

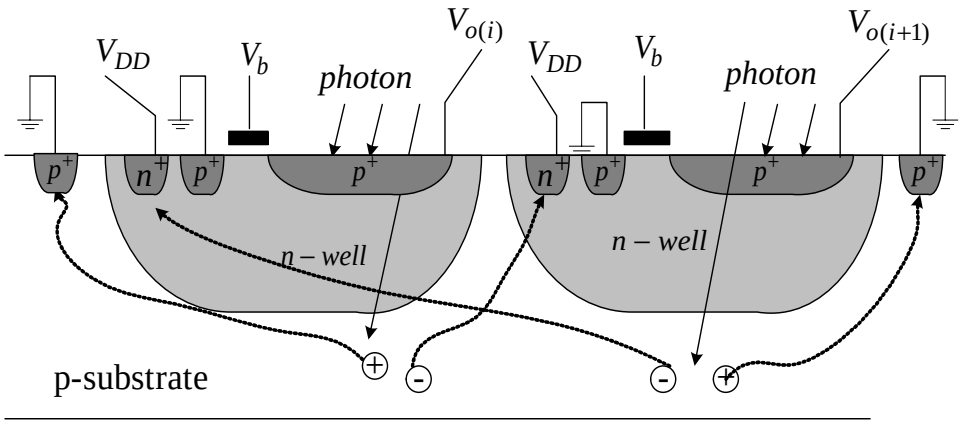


Fig. 43. The diffusion of optically generated substrate electrons/holes in our design. Note that the diffusion carriers won't affect the performances of the neighboring pixels.

strate diode. Although the quantum efficiency of $p^+/n-well$ diode can't match that of n^+/p^- substrate diode, especially for the long-wavelength photons, this architecture avoids the problems of parasitic current injecting from the substrate. As shown in Fig. 43, the photon-generated holes in the substrate will drift to Gnd and the photon-generated electrons will drift to the n-well and go to the power supply. So a pixel will never be affected by its neighboring pixels.

C. Photoreceptor Sizing and System Performance

We are going to discuss the performance of the photoreceptor from the following specs: (1) Spatial and temporal filtering; (2) Noise and Signal-to-Noise (S/N) considerations; (3) Offset consideration.

1. Spatial and Temporal Filtering

The response of the photoreceptor depends on the photocurrent, size, and parasitic capacitance. The total photocurrent I_p is proportional to Ea^2, where E is the irra-

diance and a^2 is the size of the light-collection area. The total capacitance on the photodiode is given by the sum of the junction capacitance and a contribution from parasitic due to the sensing and amplifying circuit that will be discussed in chapter V. Thus the total capacitance C is proportional to (a^2+p), where p is area-dependent [56]. The photoreceptor can be regarded as a first-order temporal filter. Its bandwidth B is inversely proportional to C and proportional to I_p. Thus

$$B \propto \frac{a^2}{a^2+p}E \tag{4.10}$$

Delbrück proved that for a given irradiance E, the bandwidth rises with the diode area, but eventually asymptotes to a constant independent of a.

For example, suppose a sharp edge moves across the sensor pixel in infinitesimal time. The irradiance of the dark and bright sides of the edge is $0.1W/m^2$ and $1W/m^2$, which generates about $\Delta V = 190mV$ shifting in V_p (Based on the simulation of Fig. 42). Capacitance of the receptor is about $200fF$. According to /refequ:photocurrent, $I_p = 100pA$. Thus the rising time t_r is given by:

$$t_r = \frac{C\Delta V}{I} = 0.4ms \tag{4.11}$$

This fits to the requirements of a real-time system whose time delays can't exceed a few tens of milliseconds [56].

Since the photoreceptor has finite size, it is a spatially low-pass filter. The effect of the square-box spatial filter is given by a sinc function $\left(sinc(x) = \frac{sin(\pi x)}{\pi x}\right)$.

2. Noise and SNR

In dc-lightning or sunlight conditions, the thermal noise is the dominant for photoreceptors. The thermal noise $V_{n(th)}$ at the photoreceptor output is caused by the amplification of shot-noise diffusion currents in the photodiode and transistors in the

photoreceptor. The mean-square thermal voltage noise of V_p is proportional to kT/C and thus given by [57]

$$V_{n(th)}^2 \propto \frac{kT}{a^2 + p} \tag{4.12}$$

According to 4.12, the thermal noise power is reduced by increasing a. With increasing a, the thermal noise from the photoreceptor becomes reduced until the input-referred thermal noise becomes dominant. Thus, increasing pixel area help improve SNR. However, there is a tradeoff that spatial filtering attenuate signal by $1/a^2$.

In practice, however, for ac-driven light sources (60 Hz line frequency in USA) that are typical for indoor lumination, optical noise is the dominant source of noise in the photoreceptor. Because the photoreceptor is a temporal low-pass filter, high frequency optical noise will be filtered. In addition, based on 4.10, smaller-size photoreceptors have better filtering effects because its bandwidth is smaller. So larger a degrades the SNR of the photoreceptor output.

3. Offset Consideration

A major drawback of logarithmic sensors is their sensitivity to device parameter variations. These variations are due to the fabrication process and they introduce an offset in the signal of each pixel. these offsets give rise to the fixed pattern noise (FPN), which appears as a time-invariant noise in every image and corrupts the image sensitivity.

In the traditional integrating sensors, FPN can be greatly reduced by correlated-double sampling [58, 51, 59]. However, there is no "reset level" in logarithmic photoreceptor. Kavadias proposed a new architecture to calibrate the offset [53]. His calibration scheme is to remove the threshold voltage offset of the transistors, which is the most critical parameter based on his measure. His scheme can't remove the

mismatch of transistor transconductance factor β.

Our focus measure processor relieves the constraints of the threshold voltage offset because it only relies on the output difference between the neighboring pixels. By careful layout design and increase the size of transistors, the mismatches ($\Delta\beta$, ΔV_{TH}) between neighboring pixels can be minimized, although the global mismatches may be still quite large. Furthermore, the experimental results given in chapter V prove that the system is insensitive to transistor mismatches.

D. Conclusions

We have shown in the physical principles of the photoreceptors, the architecture and specs of the traditional CCD and CMOS photoreceptors. We also discuss in this chapter how to build the on-chip photoreceptor and improve its performance. Given the robust, compact, and integrative features of this photoreceptor, it is more suitable for our application than those traditional and mature photoreceptors. The next chapter will discuss how to integrate this receptor with local analog processing unit and build a parallel asynchronous network to extract the range information.

CHAPTER V

SIGNAL PROCESSING CIRCUIT

In Chapter III, we saw from our study of the mathematical principle of the proposed technique that this range finding system only need light signal processing. Thus, it can be implemented with an analog signal processor. The body of this chapter is a detailed description and analysis of circuit implementation of the analog processor in the system. Analog processing is more economic in terms of silicon area and power than digital processing of comparable complexity, and thus makes higher pixel densities possible.

The essence of the analog processor is its parallel hardware implementation, where the analog processor will locally process the information captured by the photopixel. This is different from the traditional way that the outputs of the pixels are sequentially read out through a bus, converted to digital signals by A/D converters, and stored in the memory, waiting to be processed by a general digital processor. Another benefit of locally signal processing is that the signal won't be corrupted during the transmission and the signal to noise ratio (SNR) is higher.

A key component in the processor is the focus measure processor that implemented the function of the sum of modified Laplacian (SML) as discussed in Chapter III. We show that the proposed focus measure processor is ideal for local signal processing due to its compact size and linear processing region.

A. System Diagram of the Chip

Shown in Fig. 44 is the architecture of our sensor/processor chip designed for proving the proposed technique. It contains a matrix of 12 rows and 72 columns of pixels sites. The structure and response of the photopixels have been depicted in Chapter

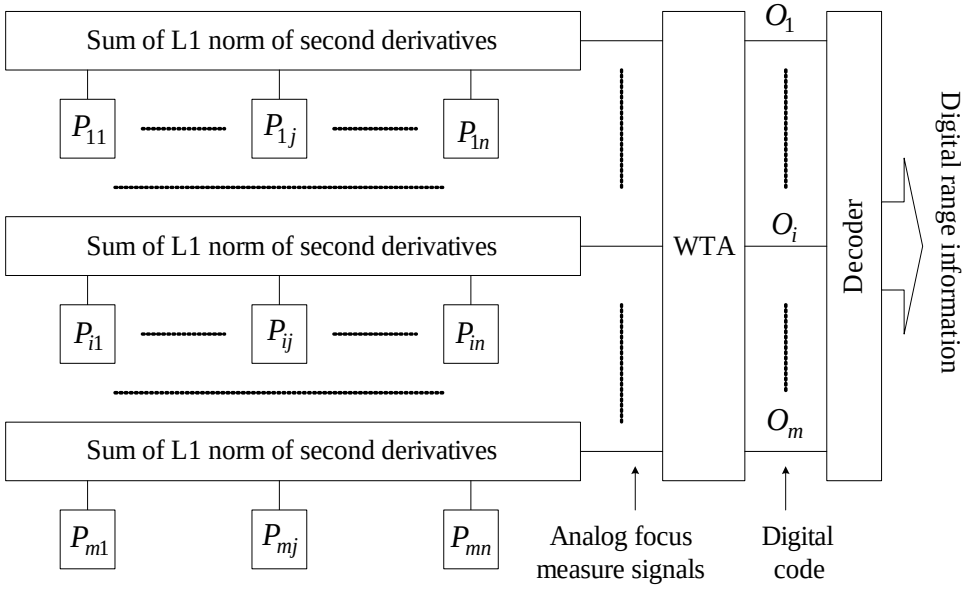


Fig. 44. The system level chip diagram. There are m pixel lines in the photosensor with n pixels in each line.

IV. The L_1 norm of second-derivative is adopted as the focus signal. Since no range mapping is intended, the signals of all 70 sites are summed along each row to create a global focus signal for each row. A Winner-take-all (WTA) processor compares all 12-row signals in parallel to identify the row of best focus.

B. Focus Measure Processor

The diagram of the row circuit with its focus measure processor/photopixel is shown in Fig. 45. Each pixel site contains a photosensor shown with a box, and a slice of focus measuring network comprising (a) two driver NMOS transistors, $M1$ and $M2$ for generating the second derivative of the photosensor output $V_{P,(i,j)}$, (b) two diode-connected NMOS transistors, for generating the absolute value of the derivative, and (c) a pair of source/sink devices for biasing the pixels with a dc current I_B. The pixels sites are also connected to two buses (bus-1 and bus-2), which are biased at

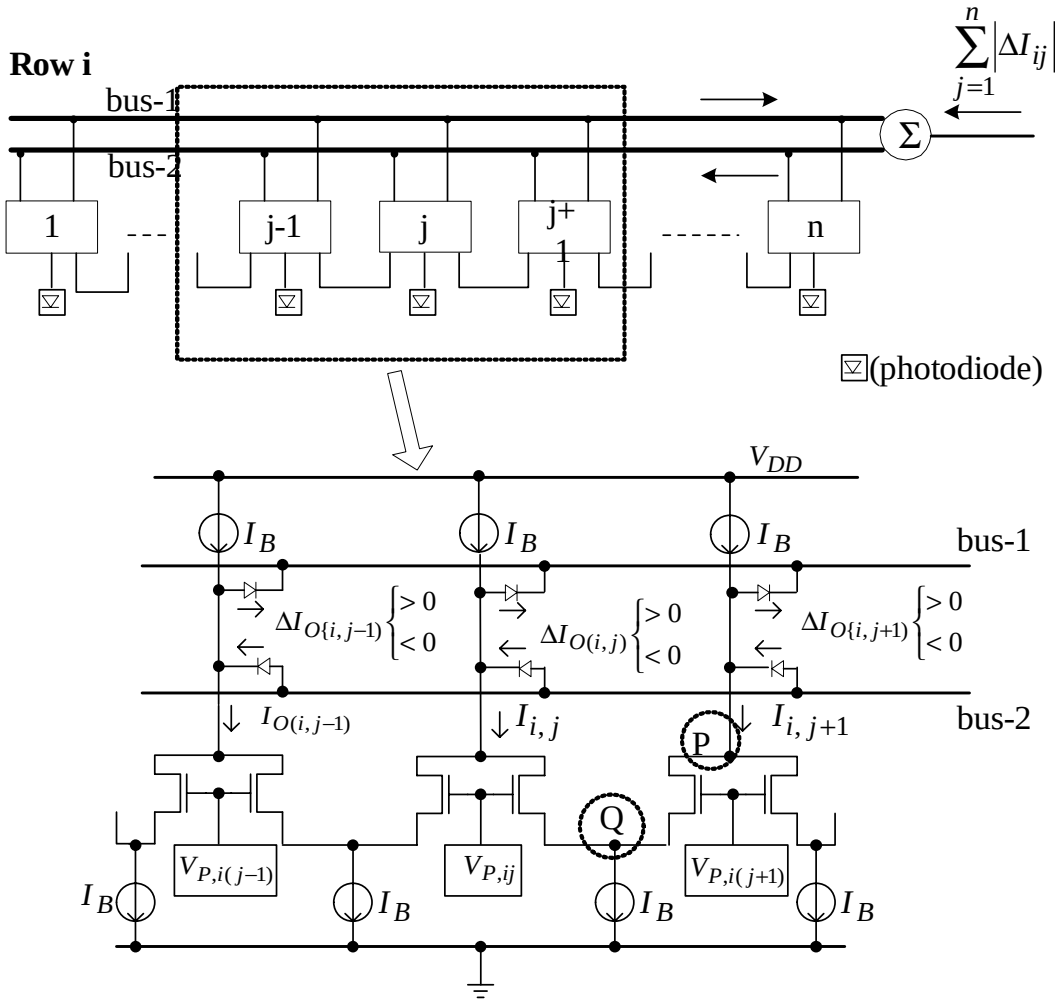


Fig. 45. Row circuitry including three neighboring pixel sites.

V_{B1} and V_{B2} respectively. The currents in bus-1 and bus-2 are to be added by a current summer that will be discussed later. Shown in Fig. 46 is the layout of a pixel site designed using AMIS $0.5\mu m$ CMOS technology, which contains a silicon area of $25\mu m * 70\mu m$.

Shown in the bottom of Fig. 45 is the circuit schematic of a 3-pixel slice of the focus processor comprising pixels $i(j-1)$, ij, and $i(j+1)$, which is also called differential differential pair (DDA). Note that each photopixel drives a pair of NMOS transistors with a common drain. The common-drain current $I_{O,ij}$ of pixel-ij can be

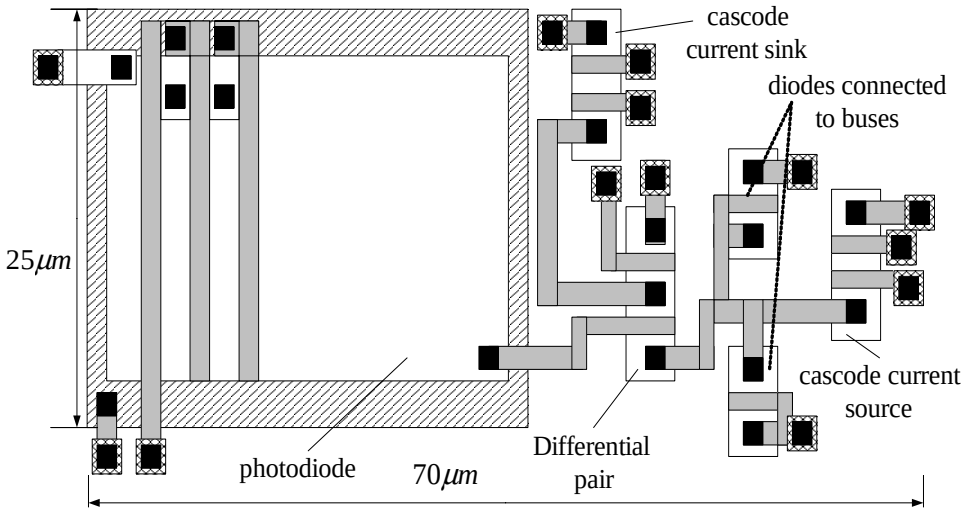


Fig. 46. The layout of a pixel site. The photodiode is formed as an extension of the source of transistor as discussed in Chapter IV. Third-level metal covers everything but the photodiode. Total area is about $25 \times 70\mu m^2$ in a $0.5\mu m$ technology.

expressed as

$$I_{O,ij} = I_B + \frac{g_m}{2}[(v_{P,ij} - v_{P,i(j+1)}) + (v_{P,ij} - v_{P,i(j-1)})], \tag{5.1}$$

where, g_m represents the MOSFET transconductance. The difference between I_B and I_i, which approximates the second spatial derivative [60, 61, 62]

$$\Delta I_{O,ij} \equiv I_B - I_{O,ij} = \frac{g_m}{2}[2v_{P,ij} - v_{P,i(j+1)} - v_{P,i(j-1)}] \tag{5.2}$$

flows out of pixel-i onto bus-1 if the second derivative is positive, and out of bus-2 into pixel-i if the second derivative is negative. These two buses are kept at virtually constant voltages by two opamps per row, and their currents are summed up to generate the sum of the absolute value of the second derivatives. Note that the diodes shown in Fig. 45 is the image sensing photodiodes.

The measurement of focus is represented by sum of modified Laplacian (SML):

$$SML = \sum_{i=2}^{n-1} |\Delta I_{O,ij}|. \tag{5.3}$$

The output current of the focus measure processor, which is the sum of bus-1 and bus-2 currents, represents the analog focus measure for each row.

1. Design Constraints

a. Effects of Nonidealities

In Fig. 45, we assume that all the current sources/sinks are ideal and the sinking/sourcing currents equal to I_B. Therefore, (5.2) can be derived from (5.1). Nevertheless, there are several factors that will affect the real value of I_B.

(1) Finite impedance of the current mirror

Since node Q in Fig. 45 is the common source of differential pair, it is virtual open and will follow the common voltage of the input gate voltages. The real sink current will be $I_{sink} = I_B + \Delta V_Q / R_Q$, where ΔV_Q is the voltage fluctuation in node Q and R_Q is the impedance of the current sink connected to Q.

The same situation will happen to the current source in the top. Whenever $\Delta I_{O,ij} \neq 0$, one of the diodes will conduct current either to bus-1 or bus-2. So the voltage in node P will change to turn on/off the diodes. The real sourcing current will be $I_{source} = I_B + \Delta V_P / R_P$, where ΔV_P is the voltage fluctuation in node P and R_P is the impedance of the current source connected to node P.

One solution to alleviate this problem is to increase the impedance of node P and Q. Therefore, cascode current sources/sinks are used in our chip (See Fig. 47).

(2) Mismatches among current sinks and current sources

As shown in Fig. 47, I_B is copied along the sensor line using current mirrors.

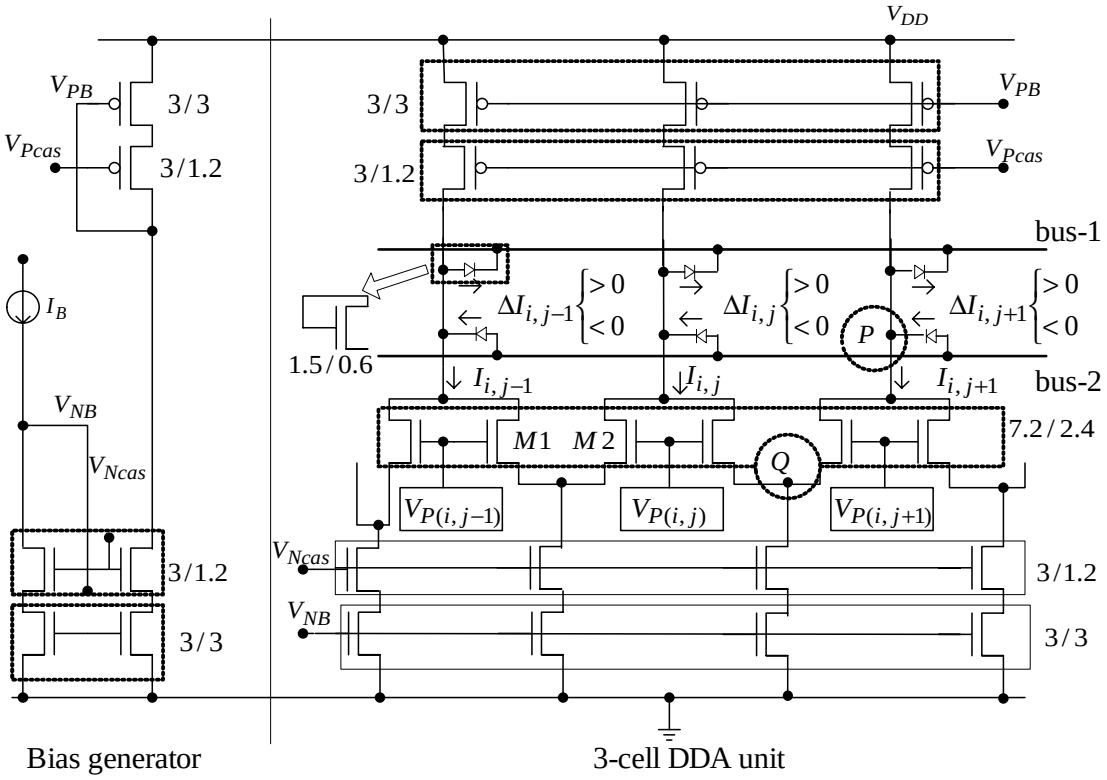


Fig. 47. Circuitry of three neighboring pixels.

The mismatches of V_{TH}, KP, and W/L will unbalanced the value of I_{source} and I_{sink}. From [63], the current mismatch can be normalized to the average value with the following equation:

$$\frac{\Delta I}{I} = \frac{\Delta(W/L)}{W/L} - 2\frac{\Delta V_{TH}}{V_{GS} - V_{TH}} \tag{5.4}$$

This result suggests that, to minimize current mismatch, the overdrive voltage must be maximized. To avoid confusion, we use I_{mis} to represent current mismatches due to the nonideality of current mirrors.

(3) Mismatches between the differential pair

The mismatches between the differential pair $M1$ and $M2$ will also generate some offset current. The offset in the input voltage due to mismatch can be expressed by [63]:

$$V_{OS,in} = \frac{V_{GS} - V_{TH}}{2}\frac{\Delta(W/L)}{W/L} - \Delta V_{TH} \tag{5.5}$$

The current offset due to this voltage offset is $I_{OS} = g_m V_{OS,in}$.

To minimize the mismatches between the current mirrors and differential pair, careful layout is important. There are three ways to improve matching [64]: (a) Common-centroid layout helps match errors caused by gradient effects. (b) Making larger objects out of several unit-sized components connected together. (c) Matching the boundary conditions around all components to be matched.

(4) Nonideality of diodes

In digital CMOS process, the diodes are usually implemented using diode-connected NMOS.

Due to body effect, V_{TH} of the diode-connected NMOS is large ($> 1V$). Furthermore, since bus-1 and bus-2 are biased at fixed voltages, to turn on/off the diodes, the voltage changing in node P is quite large. This voltage changing in turn affects

the current sources. Because this, $V_{GS}-V_{TH}$ of the diode-connected NMOS shouldn't be too negative to avoid large voltage changing in node P.

On the other hand, MOSFET transistors have leakage/subthreshold currents in the subthreshold region, which can be represented by $I_{sub} \approx I_0 exp(V_{GS}/(nV_T))$ [55], where $n \approx 1.5$ and $V_T = kT/q$. I_{sub} will flow from bus-1 to bus-2 and be doubled by the current summer shown in Fig. 45. From the simulation, the average value of I_{sub} is in the range of $1 - 10nA$.

If all those effects are taken into consideration, 5.2 should be modified as:

$$\Delta I_{O,ij} = \frac{g_m}{2}[2v_{P,ij} - v_{P,i(j+1)} - v_{P,i(j-1)}] + \Delta V_P/R_P - \Delta V_Q/R_Q + 2I_{sub} + I_{OS} + I_{mis}, \quad (5.6)$$

where, $\Delta V_P/R_P - \Delta V_Q/R_Q + 2I_{sub} + I_{OS} + I_{mis}$ can be processed as noise No_{ij} associated with the unit processor ij, whose distribution can be represented by $N(\mu, \sigma)$. Ideally, $\mu = 0, \sigma = 0$. σ is mainly decided by the mismatches between pixelsites.

Total output current of the current summer in row i can be represented by:

$$I_{O,ij} = \sum_{j=2}^{n-1} |\Delta I_{O,ij}| + \sum_{j=2}^{n-1} |No_{ij}|. \quad (5.7)$$

The first item in 5.7 is the focus measure, and the second item is the integrated noise. Since the noise distribution in each processor unit is uniform, $\sum_{j=2}^{n-1} No_{ij}$ will be a normal distribution with μ and σ ($N((n-2)\mu, \sqrt{N}\sigma)$. If the number of cells in a row n is fixed, $\sum_{j=2}^{n-1} No_{ij}$ is close to a constant $C = (n-2)\mu$, where C is an unknown value. Here we can see that the mean value of noise C will increase with the number of the cells, which means the noise level will linearly increase as the number of pixelsites increases. Fortunately, C is the common mode current and can be easily removed by the WTA circuit, which only detects the difference between the input currents/voltages. Therefore, we only need to worry about the stand deviation of the row noise current, which is given by $\sqrt{N}\sigma$. As we can see, the stand deviation of the

noise current increase much slower as the row pixelsites number increase. We can also expected that the row focus measure current has linear relationship with the number of the pixels, since the longer the pixel row, the more edges/textures it will detects. In this way, increasing the number of pixelsites in a row helps to increase SNR.

Based on the discussions above, the matching between pixelsites mainly decide the noise level in our system. Therefore, the importantance of careful layout design should never be forgot.

b. Linearity of DDA

The DDA used to extracted local focus measure converts the gradient of photovoltages into current. This conversion should be linear in a wide range. The linear range of DDA is decided by the linear range of the differential pairs since it simply adds two the differential current.

The differential current generated by M1–M2 is given by [63]:

$$I_{dff} \approx K_P \frac{W}{L} V_{in} \sqrt{4(V_{GS-TH})^2 - V_{in}^2},$$

where V_{in} is the input voltage difference.

If $|V_{in}| \ll V_{GS} - V_{TH}$, then

$$I_{dff} \approx K_P \frac{W}{L} V_{in} (V_{GS-TH}) [1 - \frac{V_{in}^2}{8(V_{GS-TH})^2}] \tag{5.8}$$

(5.8) indicates that increasing $V_{GS} - V_{TH}$ can improve the linear range. There are two ways to increase $V_{GS} - V_{TH}$: (1) Increase the tail current; (2) Decrease the W/L ratio.

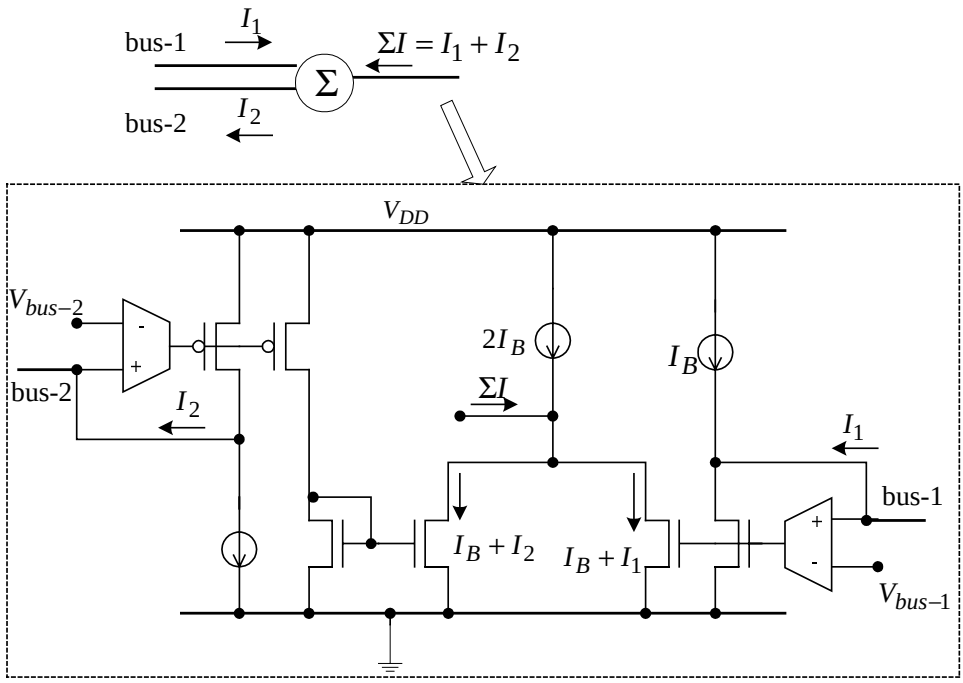


Fig. 48. Block diagram of the current-summer, which add the currents flowing the bus-1 and bus-2.

2. Current Summer

The currents flowing in bus-1 and bus-2 will be summed up by a current-summer connected to each row. Shown in Fig. 48 and Fig. 49 are the block diagram and transistor-level circuit of the current-summer. Its output current ΣI is the row focus measure (sum of modified Laplacian).

3. Experimental results of focus measure processor

In the first prototype chip, we deliberately separated one row to test the functionality of the focus measurement processor. Fig. 50 shows its circuit diagram. The only difference is that this row has only three processor units whose inputs come from outer signal source rather than the photopixels and the output current is directly connected to a transimpedance amplifier. This provides us a way to control the inputs to the

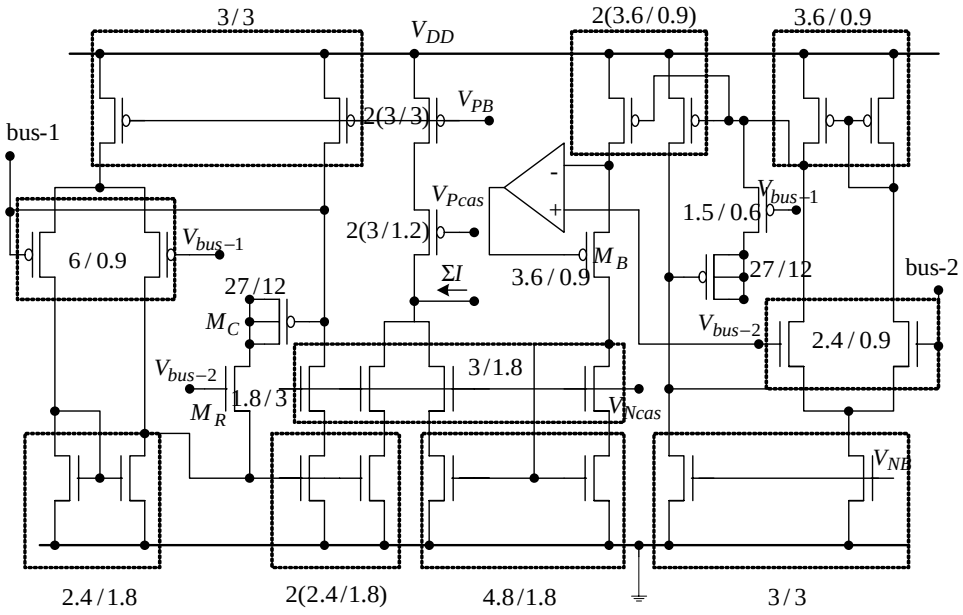


Fig. 49. Transistor level current-summer design.

focus measurement processor. The output of the focus measure processor is given by

$$V_o = V_B + g_m|(2V_2 - V_1 - V_3)|R \tag{5.9}$$

Given in Fig. 51 is the oscilloscope traces of the input and output voltages measured when $V_1 = V_3 = 3.5V$, $I_B = 5\mu A$, and V_2 is a triangular signal with $f = 1kHz$, $V_{offset} = 3.5V$, and $V_{pp} = 200mV$.

Experimental results of the 3-cell focus measurement processor are summarized in Table VI. Vi_{dff} is the differential input voltage, Vi_{com} is the common mode input voltage, and f_{Vi} is the frequency of the input signals. Table VI verifies the linear range and speed of our focus measure processor.

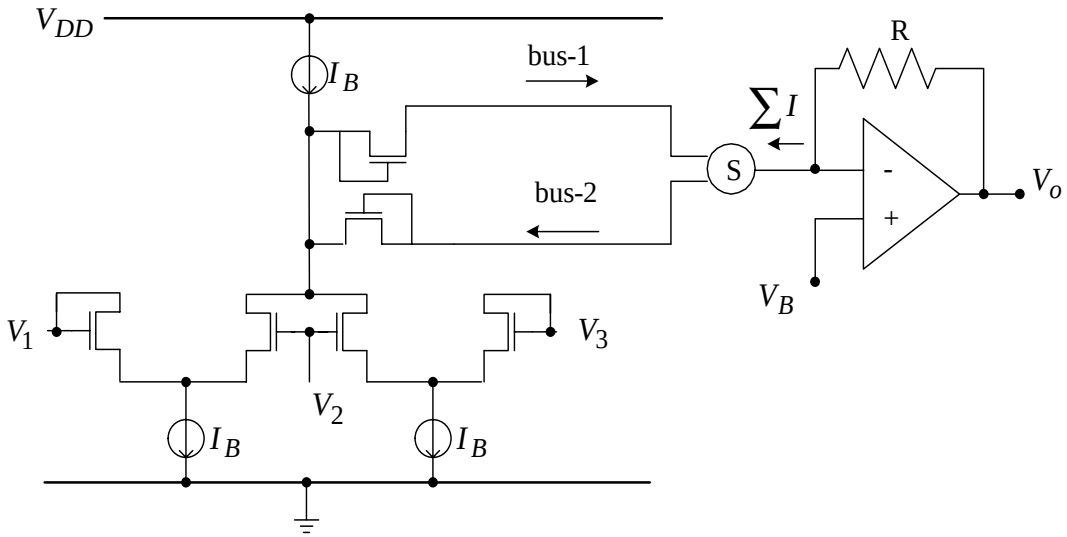


Fig. 50. Circuit diagram of the test row.

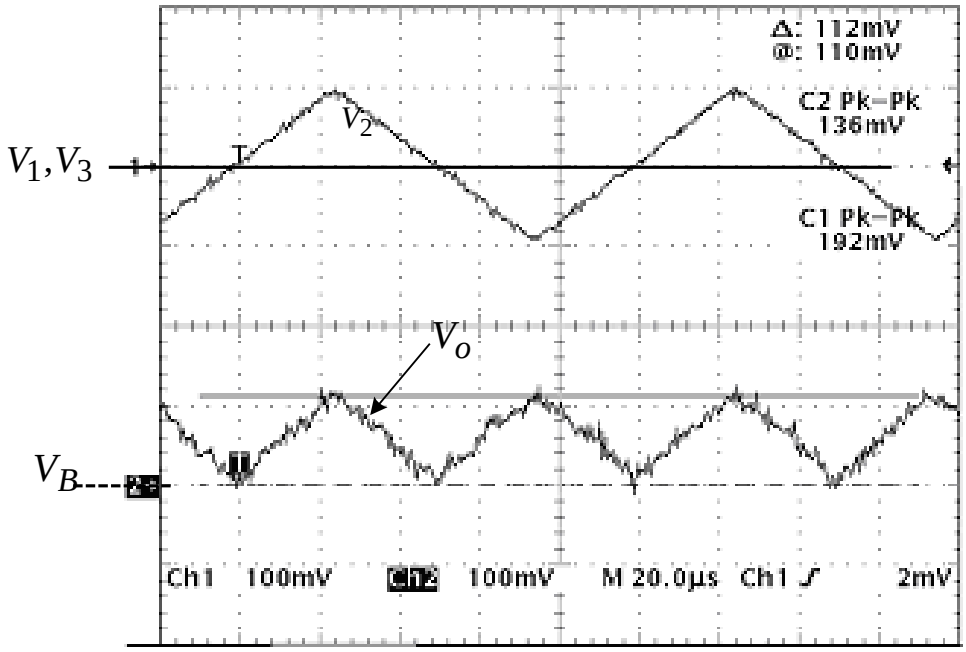


Fig. 51. Output waveform of V_2 and V_o. Note when $V_1 = V_2 = V_3$, $\Sigma I = 0$ and $V_o = V_B$.

Table VI. Experimental results for the 3-cell focus measurement processor

Maximum linear Vi_{dff}	150 mV
Vi_{com} Range	$(2.7V, 4.2V)$
Maximum f_{Vi}	up to $1MHz$

C. Winner-take-all Circuit

The winner-take-all (WTA) circuit, which chooses a winner from a group of input signals, is a basic and important analog building block. The inputs of WTA can be voltage or current. Considering the current-summer output, our system uses a current-mode (CM) WTA. A current-mode MOS implementation of the WTA function was first introduced by Lazzaro et al. [65], and has been used by others in weak inversion [66] and strong inversion [67].

1. Lazzaro WTA

Fig. 52 shows the basic architecture of Lazzaro WTA. Each cell contains two NMOS transistors, $M1_i$ and $M2_i$. Its input current is I_i and output voltage is V_i. The circuit operates by choosing the maximum input current I_m and broadcasting its value as a winner onto the global line V_g. Suppose in the beginning all the M1 transistors sink the same current, because all the M1 transistors share the same gate to source voltage V_g, the voltage of drain of $M1s$ are equal. In time t, suppose cell m receives larger current I_m ($I_m > I_i$ for all $i \neq m$), drain voltage of M1 in cell m increase, and due to the effect of source follower M2, voltage of V_g increases. The increases of V_g reduces the gate to source voltage of all $M2_i$ transistors, hence decreasing the current through every $M2_i$ transistors. As the summation of current through all $M2$

transistors equals to I_B, current through $M2_m$ increases, this further increases the voltage of V_m. As the results of the competition, the cell that received highest input current has the highest output voltage. If the difference between I_m and any other I_i is large enough, only $M1_m$ and $M2_m$ stay in the saturation region, all $M1_i$ enter linear region and $M2_i$ are cut off.

The Lazzaro WTA circuit accomplishes WTA function by only inhibitory competition among the cells. This means that a small/large difference between the winner's input and other cells' inputs results in small/large difference between the winner's output and other cells' output. This drawback harms our system. For example, under low illumination, the focus measure current of winner row may be quite small compared with I_B, which corresponds to the small voltage difference between the winner O_{win} and other O_i in Fig. 44.

A $12-cell$ WTA using the traditional Lazzaro circuit has been simulated using BSIM model. All the $M2$ transistors have the same size $(W/L)_2 = 3\mu m/1.5\mu m$ and all the $M1$ transistors have the same size $(W/L)_1 = 1.5\mu m/4.5\mu m$. The current sources are cascode PMOS transistors. Given in Fig. 53 are the simulation results. The input setting is: $I_{1-5,8-12} = 10\mu A$, $I_6 = 10+1\mu A$, $I_7 = 10+0.7\mu A$. Because I_6 is the largest input current, cell 6 is supposed to be the only winner. However, as shown in Fig. 53, the difference between the winner and the closest loser (3% difference) is less than $0.5V$, which indicates that this Lazzaro circuit can't distinctly separate the winner from the losers for a small input difference. Also as shown in Fig. 53, the Lazzaro circuit needs 10 percent input difference to clearly separate the winner from the losers.

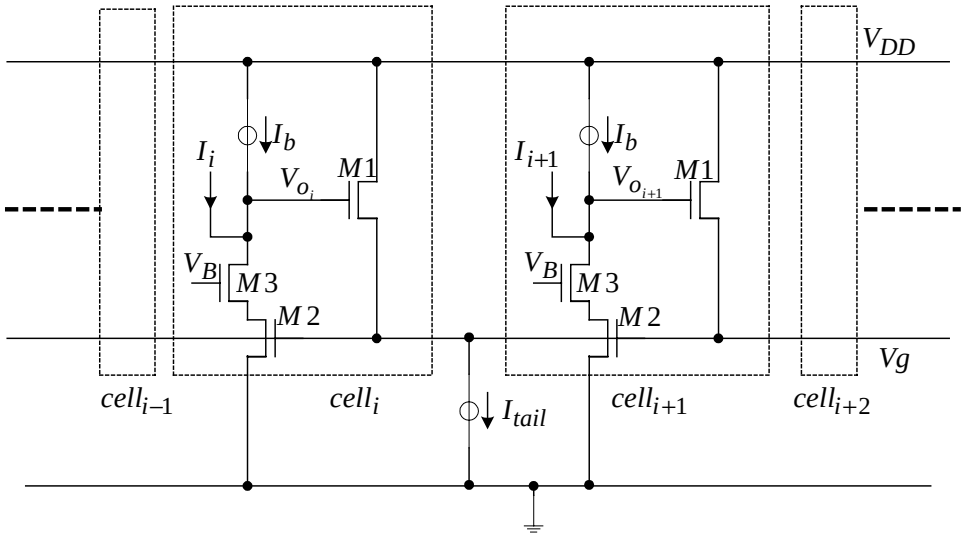


Fig. 52. Lazzaro WTA circuit.

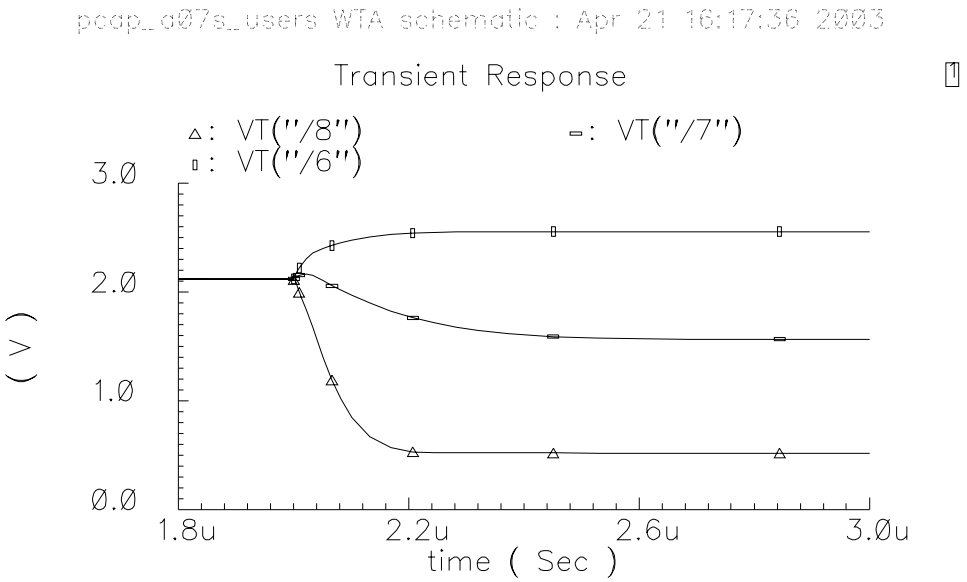


Fig. 53. Transient response of the Lazzaro circuit with 3% and 10% input difference. $VT('6')$: Winner's output ($I_6 = 11\mu A$). $VT('7')$: $Loser1$'s output ($I_7 = 10.7\mu A$). $VT('8')$: $Loser2$'s output ($I_8 = 10\mu A$). Input currents for all other cells except for $cell6$ and $cell7$ are set to $10\mu A$.

.

2. Modified Lazzaro WTA

Starzyk and Fang [68] described a modified Lazzaro circuit containing both excitary and inhibitory feedback, which greatly increases the sensitivity without losing the speed. However, the positive feedback causes the circuit to be stuck in a stable stage after competition is completed, so the circuit needs to be reset after every competition. This requirement makes this modified version unsuitable for our design since the WTA in our system continuously monitors the output current of each row. The competition between rows is real-time and can't be stopped.

The next section will discuss some design constraints of a current mode WTA circuit.

3. Design Constraints

Besides the noise factor, the limitations of a WTA circuit include at least: (1) Random resolution limitation which stems from device mismatch. This limitation depends on physical and geometrical parameter deviation; (2) Systematic resolution limitation. This factor can be ascribed to the finite open-loop gain of the gain stage.

Now I'm going to analyze the two aspects:

1. Mismatch factor

First we can see that mismatch between M1s in the cell is not important. The output voltage V_{oi} of each WTA cell i is given by

$$V_{oi} = V_g + V_{THi} + \sqrt{2I_{M1}/(K_P W/L)}. \tag{5.10}$$

From (5.10), 1% mismatch in V_{TH} and transistor size between M2s will bring very small mismatch in V_{oi}. Second, for transistor M2, we have [63]

$$\left(\frac{\Delta I}{I}\right)^2 = \left(\frac{\Delta(W/L)}{W/L}\right)^2 + 4\left(\frac{\Delta V_{TH}}{V_{GS} - V_{TH}}\right)^2 \tag{5.11}$$

From (5.11), assume $V_{GS} \approx 2V_{TH}$ and there is 1 percent mismatch in transistor size and threshold voltage between M1s, the upper limit of WTA resolution approximately equals to 2 percent of I_b.

2. Gain factor

The differential voltage between evaluation nodes of the winning cell and the losers can be regarded as the voltage drop across r_o by the current difference. We have $\Delta V_{oi} = \Delta I_{in} r_o$, where r_o is the impedance measured in the cell output. If ΔV_{diff} is the minimum output difference for WTA to pick up a winner m, we have

$$V_{om} - V_{oi} > \Delta V_{diff} \text{ for all } i \neq m$$

.

The resolution of WTA $I_{resolution}$ can be found by

$$I_{resolution} = \frac{\Delta V_{diff}}{r_o}. \tag{5.12}$$

To achieve high resolution, r_o should be as large as possible. Ideally, the resolution limit can be zero if $r_o = /infty$.

Our system uses a modified WTA proposed by Sekerkiran and Çilingiroğlu [66] to improve the WTA resolution. As shown in Fig. 54, the modified structure uses cascode stage to increase the value of r_o. In such a configuration output resistance of the part sinking the inhibitory current is given by:

$$r_o = Rds_{M2} Rds_{M3} gm_{M3}, \tag{5.13}$$

where Rds_{M2}, Rds_{M3} are drain-to-source resistance of $M2$ and $M3$, gm_{M3} is the transconductance of $M3$. This cascode structure boosts the output resistance which is equals to Rds_{M2} in the original Lazzaro circuit by a factor of $gm_{M3} Rds_{M3}$, resulting in much better resolution.

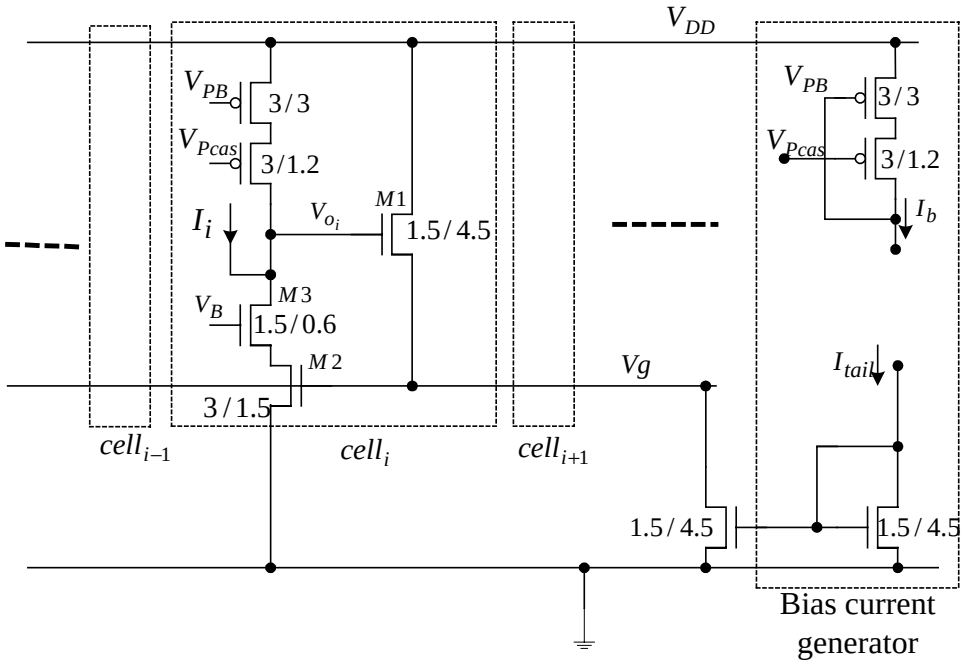


Fig. 54. Modified Lazzaro WTA circuit.

Shown in Fig. 55 is the simulation result of the modified WTA circuit using the same simulation setup. Because I_6 is the largest input, cell 6 is supposed to be the only winner. As shown in Fig. 55, the difference between the winner and the closest loser (3% difference) is larger than $2.5V$, which indicates that this modified Lazzaro circuit has much better resolution than the original circuit that needs 10% input difference to reach the same separation level.

D. Readout Circuit

As shown in Fig. 56, the first prototype also includes a network of clocked switches by which the global focus currents of all rows can be disconnected from WTA inputs, and diverted to an output pad sequentially. This enables us to bypass the WTA, and observe the analog behavior of the pixel matrix.

Shown in Fig. 57 is the readout stage, where nodes In_1, In_2, ..., In_{12} are con-

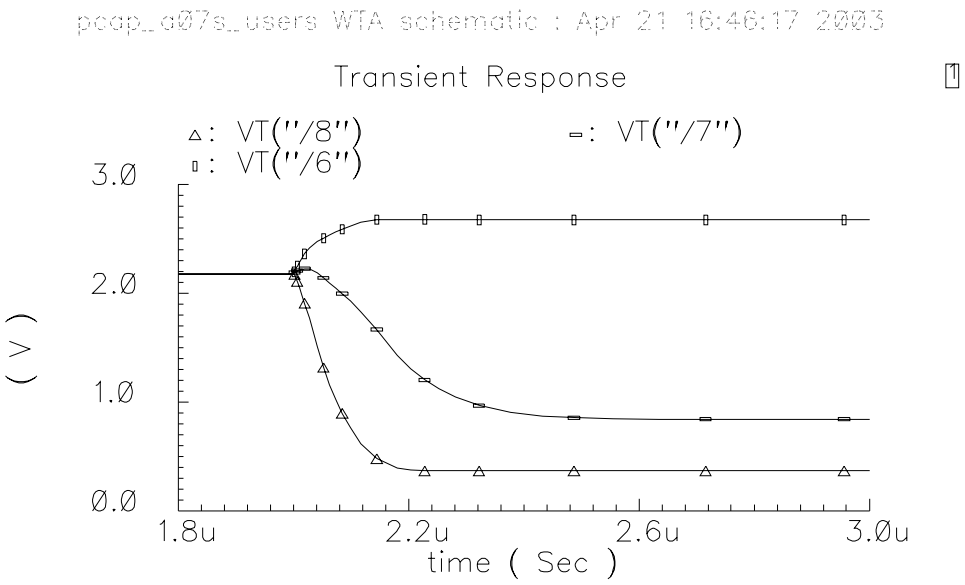


Fig. 55. Transient response of the modified circuit with 3% and 10% input difference. $VT('6')$: Winner's output ($I_6 = 11\mu A$). $VT('7')$: $Loser1$'s output ($I_7 = 10.7\mu A$). $VT('8')$: $Loser2$'s output ($I_8 = 10\mu A$). Input currents for all other cells except for $cell6$ and $cell7$ are set to $10\mu A$.

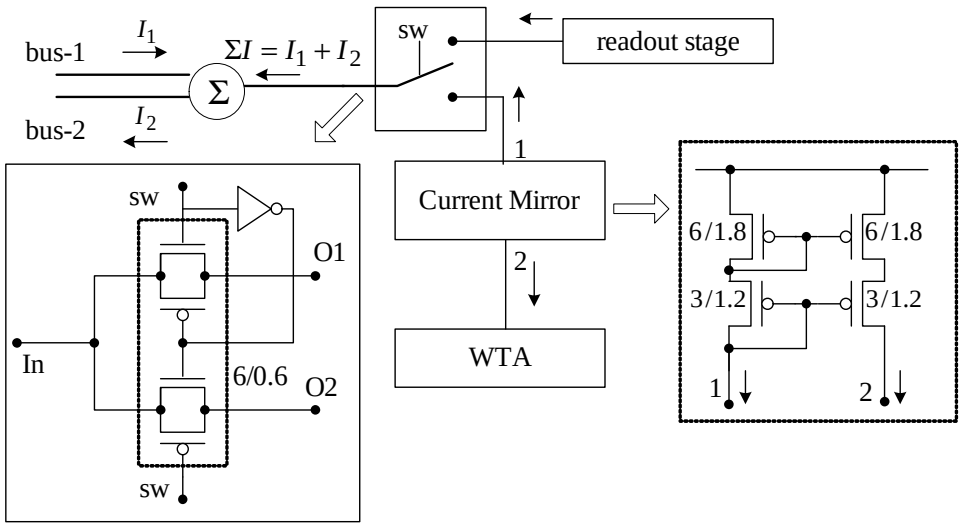


Fig. 56. Block diagram of clocked switch network.

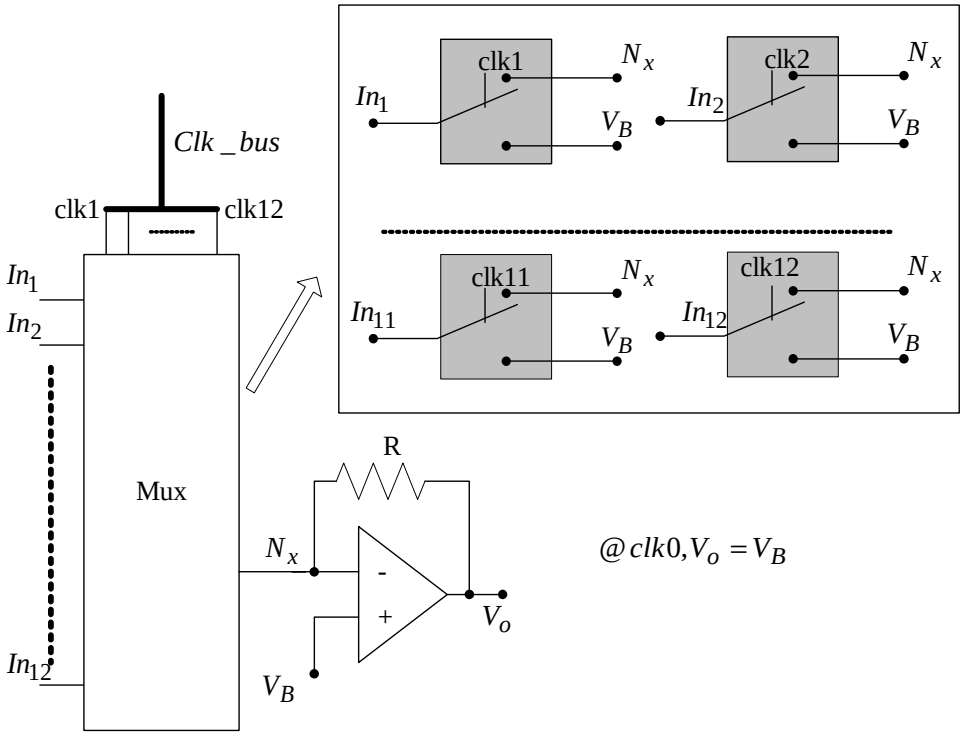


Fig. 57. Block diagram of readout stage.

nected to the outputs of the row current-summer$(1-12)$, respectively. The focus measure current from each row is switched to node N_x or V_B by a clocked switch. Node N_x is connected to the input of a transimpedance amplifier.

For any switch i in Fig. 57, if clock signal $clki$ is high, node In_i is connected to node N_x. Otherwise, node In_i is switched to node N_B that is biased at V_B. Since there is virtual short in the input nodes of output Opamp, the voltage of N_x equals to V_B. Therefore, no matter $clki$ is high or low, node In_i is always biased at V_B. This avoids the disturbance to the current-summer and promises the high speed of the readout circuit.

When node In_i is connected to N_B, the focus measure current I_i was provided by node N_B, which is connected to a voltage source. Otherwise, I_i is provided by the transimpedance amplifier and creates a voltage drop across R. The voltage measured

in node N_o is

$$V_{N_o} = V_B + I_i R. \tag{5.14}$$

Since the readout block will read the output current from $row1$ to $row12$ repeatedly, we need a reset clock to mark the beginning of each reading circle. The reset clock is $clk0$, which is not shown in Fig. 57. When clk0 is high, $clk1 - clk12$ are low and node N_x isn't connected to any N_i, since clk0-clk12 are nonoverlapping clocks. Therefore at this moment, $V_{N_o} = V_B$. Techniques to generate nonoverlapping clocks ($clk0$-$clk12$) from a single clock Msr_clk may be found in literature [64, 69]. Note that $clk0 - clk12$ must be nonoverlapping clocks. For example, if there exists a overlapping time between clk1 and clk2, row1 and row2 will be shorted together and this short circuit will report wrong range information.

Shown in Fig. 58 is the clock timing diagram needed to drive the switching network.

1. Clock Generator

The first step is to generate clocks signal ϕ_{0E}, ..., ϕ_{12E} using the circuit shown in Fig. 59. This circuit includes a sequential circuit consisting of a shift-register, and a combinational circuit consisting of 3 logic gates, which provides $\phi_{0E} = \overline{\phi_{1E} + ... + \phi_{12E}}$. When the circuit is powered on, outputs of DFFs may be '1' or '0'. In this case, ϕ_{0E} is '0'. The shifter register will set all the D-FF outputs to '0' after 12 Msr_clk cycles. Only at that moment can ϕ_{0E} be set to '1'. Therefore, after 12 Msr_clk cycles, there is only one D-FF output can be '1' and the clock generator starts to create correct clock signals.

The method of generating n nonoverlapping clocks is shown in Fig. 60. That circuit uses the feedback from adjacent NOR gates and inverter-delayed buffers to

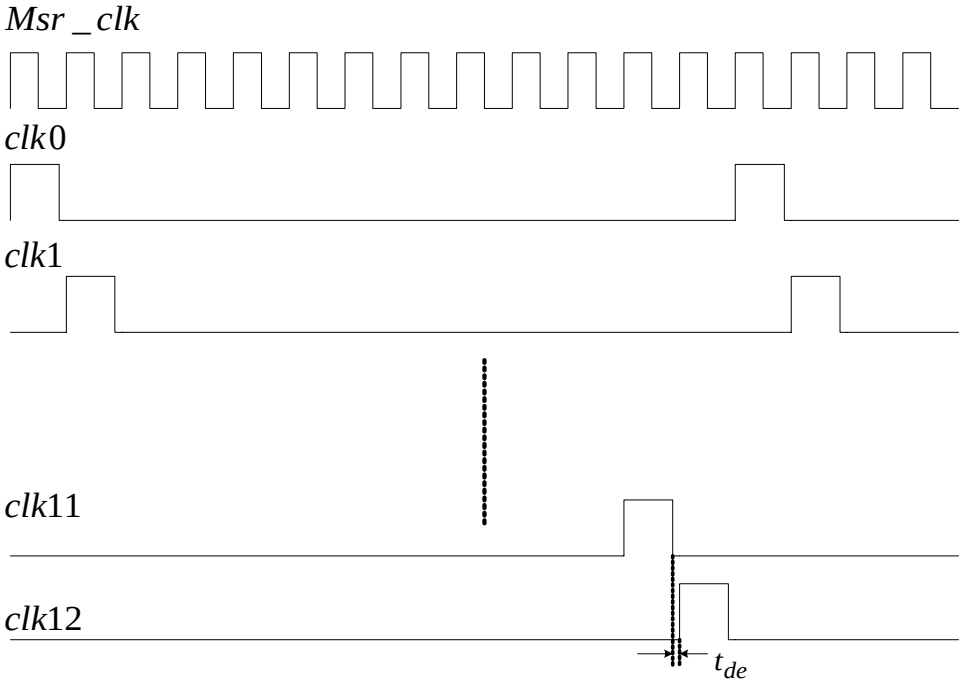


Fig. 58. Timing diagram of 13-phase nonoverlapping clocks. Msr_clk is the master clock signal to generate *clk*0-*clk*12.

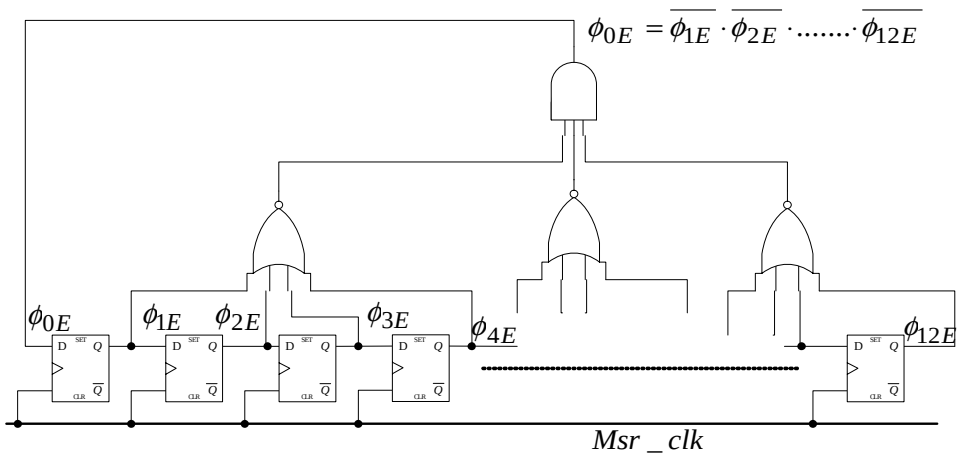


Fig. 59. Circuit diagram to generate 13-phase clocks. Msr_Clk is the master clock signal to drive the shift register.

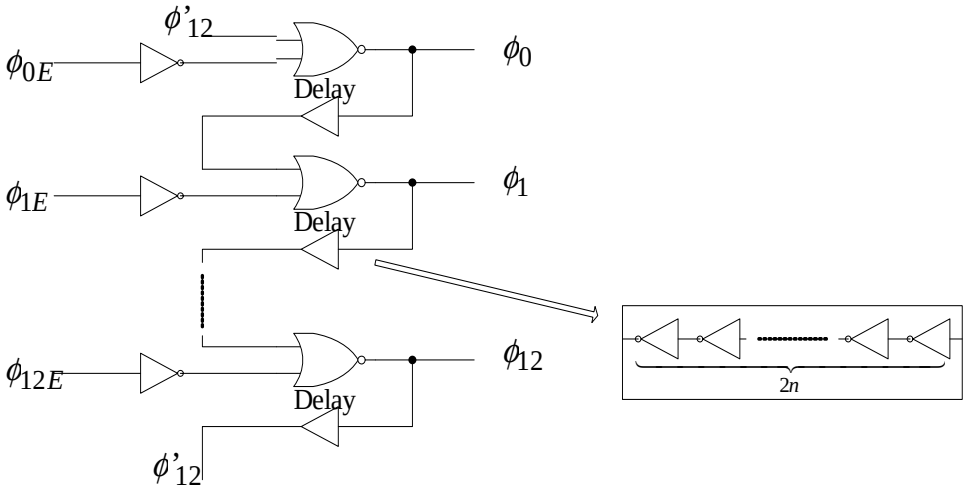


Fig. 60. Method to generate n nonoverlapping clocks.

ensure nonoverlapping. Neglecting the delay of the NOR gate, the nonoverlapping time can be adjusted by the delay of the inverter-buffers. The rise and fall time of the long-channel inverter output is given by

$$t_{fall} \approx \frac{C_{load} T_{ox} L_n V_{DD}}{\mu_n \epsilon_{ox} (V_{DD} - V_{Tn})^2} \tag{5.15}$$

$$t_{rise} \approx \frac{C_{load} T_{ox} L_p V_{DD}}{\mu_p \epsilon_{ox} (V_{DD} - V_{Tp})^2} \tag{5.16}$$

Therefore, the total delay of this chain of $2n$ inverters is given by

$$t_{delay} = \sum_{i=1} n[t_{fall(i)} + t_{rise(i)}]. \tag{5.17}$$

2. Experimental Results

Shown in Fig. 61 is the oscilloscope traces of the clock signals, where Msr_clk is the input clock signal, clk1 and clk2 are nonoverlapping clocks generated by the clock generator.

Shown in Fig. 62 is the oscilloscope traces of the row.

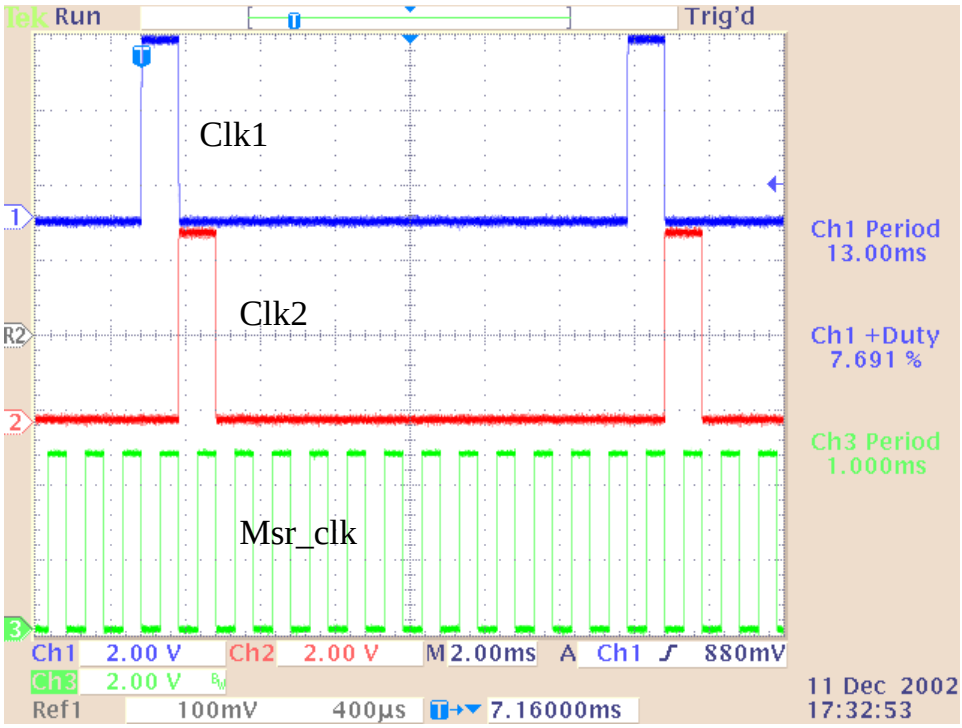


Fig. 61. Oscilloscope traces of the generated clocks together with the input clock.

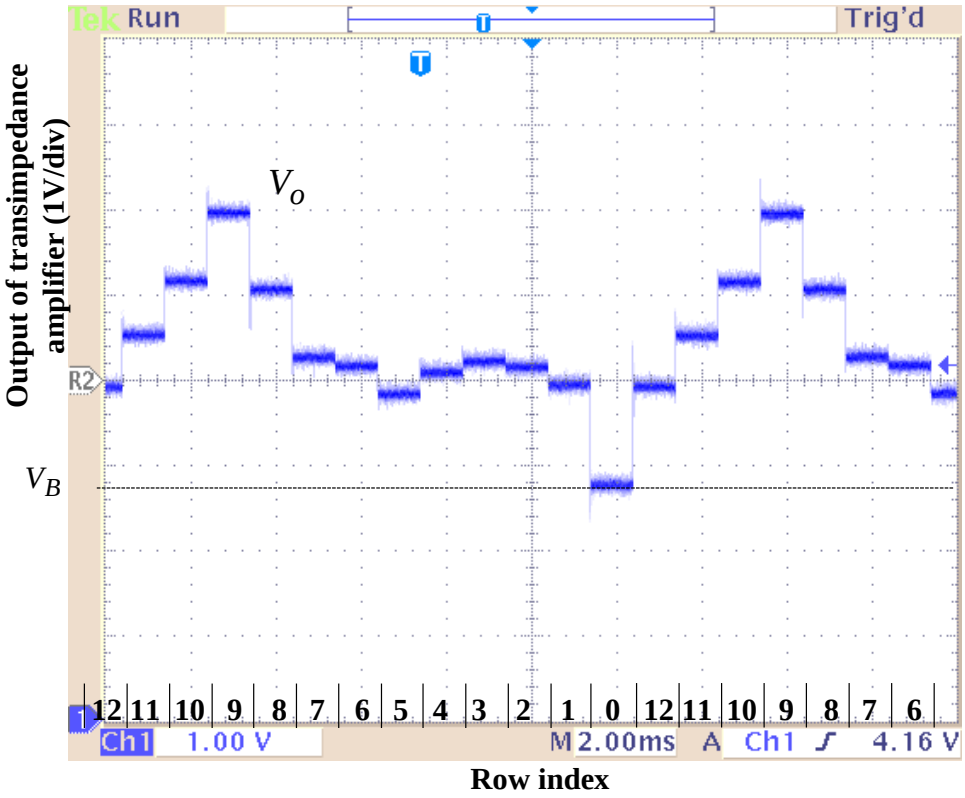


Fig. 62. Oscilloscope traces of the output of readout stage. The readout voltages has been converted to the global focus currents by $I = V/R$.

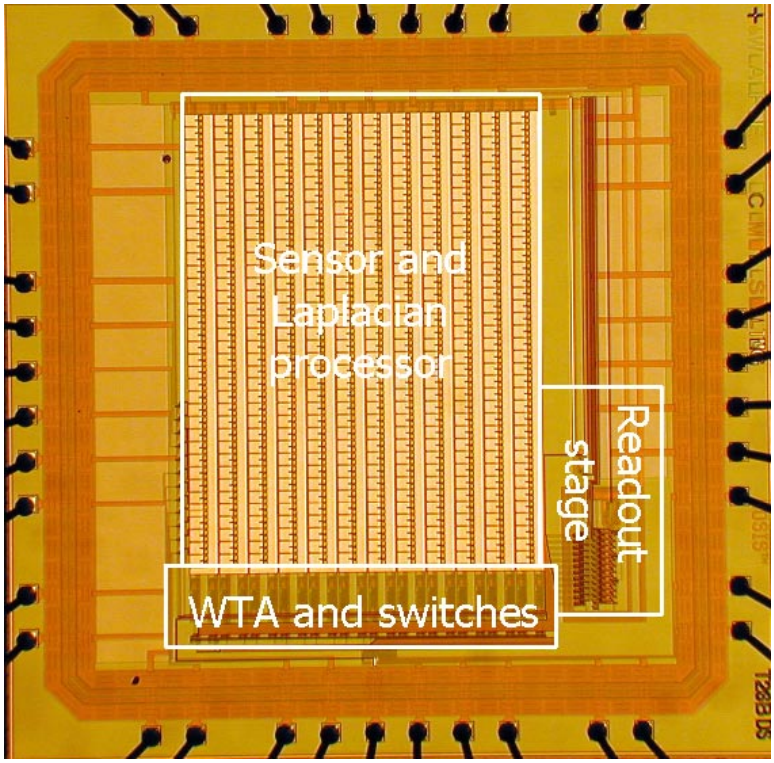


Fig. 63. Photomicrograph of the sensor/processor chip fabricated in $0.5-\mu m$ CMOS. The active area of the chip is $1.80 \times 1.44 mm^2$.

E. Experimental Results of the First Prototype

In order to verify the functionality and the performance of the proposed range-finder microsystem, this real-time range finder chip has been fabricated in a three-metal $0.5\mu m$ CMOS process. Fig. 63 shows a die photo of the sensor/processor chip that is built based on the architecture shown in Fig. 44. Although pixel size is $25 \times 70\mu m^2$, the row pitch is set to a larger $120\mu m$ in order to avoid potential selectivity problems in the first prototype. When operated from a $5V$ supply, the system dissipates $30mW$. The chip uses a DIP40 package.

Table VII. The nominal range in centimeters sensed by each row in the experimental setup

Row index	1	2	3	4	5	6	7	8	9	10	11	12
Range (cm)	543	459	397	350	312	282	258	237	219	204	191	179

1. Testing Setup

In order to carefully adjust the position of chip so that the sensor plane intersects the focal point F, a camera body is designed and fabricated to fit the chip and its board. The printed board is a two-sided, copper-clad board. The transimpedance amplifier shown in Fig. 57 that will convert the focus measure current to voltage is a commercial opamp $lm248$. V_B is biased at $2.7V$. This prototype doesn't need any clock signal; however, since the WTA outputs won't give us much information of the each row, the outputs of the row current-summers are diverted to an output pad sequentially by the readout circuit. The $1kHz$ square wave is the input clock signal of the on-chip clock generator, which provides the readout clocks (clk0-clk12). The clock generated has been tested with input clock frequency up to $100kHz$. Since the potential applications of this system are mainly for collision-avoidance, there is no need to further increase the frequency of the digital clocks. The lens used here is a Pentax $50mm$ SLR lens whose f-stop can be adjusted from 1.4 to 8. The sensor plane is tilted by 45^o. The environmental setup is shown in Fig. 64. Given in Fig. 65 are two object patterns used to test the prototype. Testing board1 is a $5mm \times 5mm$ checkerboard as shown in Fig. 65(a), and testing board2 is the color cover of a magazine shown in Fig. 65(b).

The nominal range expected to be sensed by each row for these setting and for the row pitch $120\mu m$, are calculated and given in Table VII.

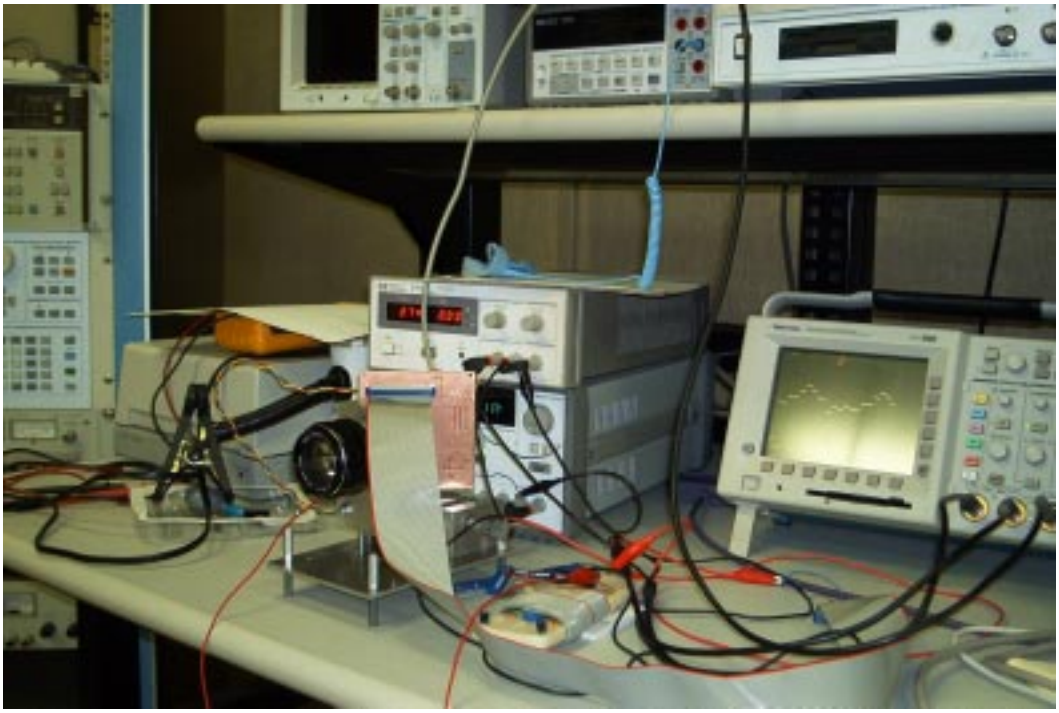
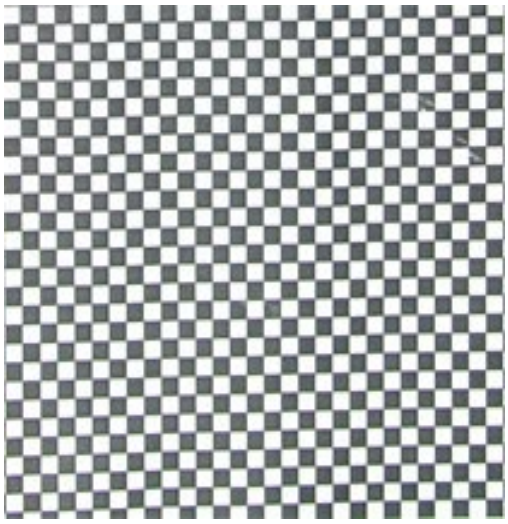


Fig. 64. The testing environmental setup of the range finder prototype.



(a)

(b)

Fig. 65. Two testing object patterns. (a) $5mm \times 5mm$ checkerboard. (b) The color cover of a magazine.

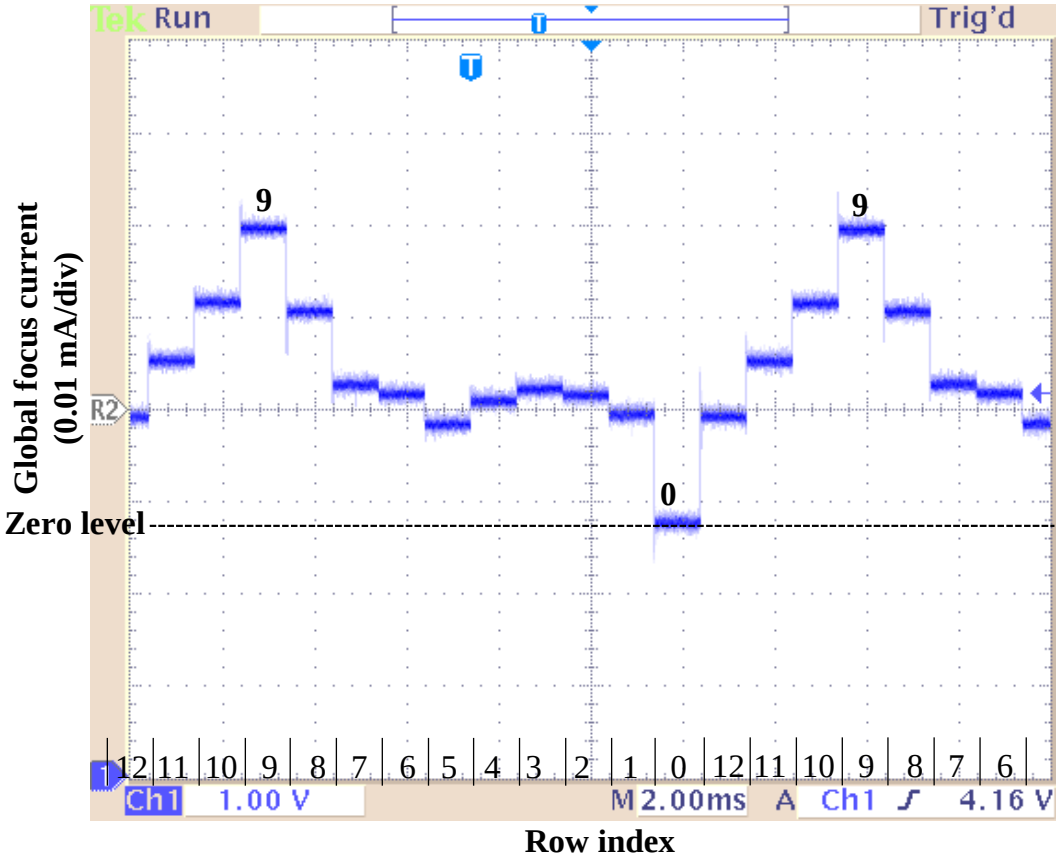


Fig. 66. Oscilloscope traces of global focus currents measured for all 12 rows when testing board1 was put 220cm away.

2. Experimental Results of the Prototype

Given in Fig. 66 and Fig. 67 are the focus measure currents for all 12 rows when object when testing board1 was put 220cm and 450cm away, respectively. The winner row is row9 and row2 respectively, which clearly verify the proposed technique. In both figures, the output currents difference between the winner and its closest competition lines are more than 10μA or 30%, which is much larger than the resolution of the WTA (3% based on the simulation). Note that in this first prototype, we have just 72 pixels in each row. We can easily build more cells in each row to further increase the output current of the winner row.

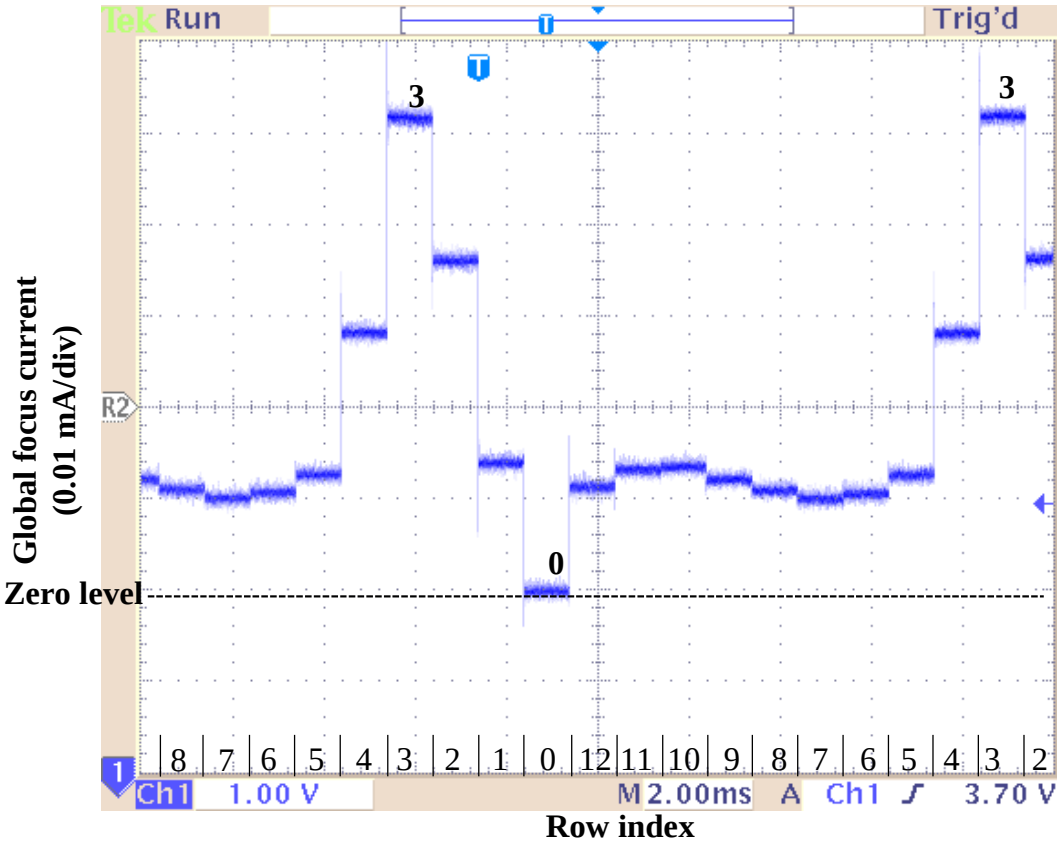


Fig. 67. Oscilloscope traces of global focus currents measured for all 12 rows when testing board1 was put 440cm away.

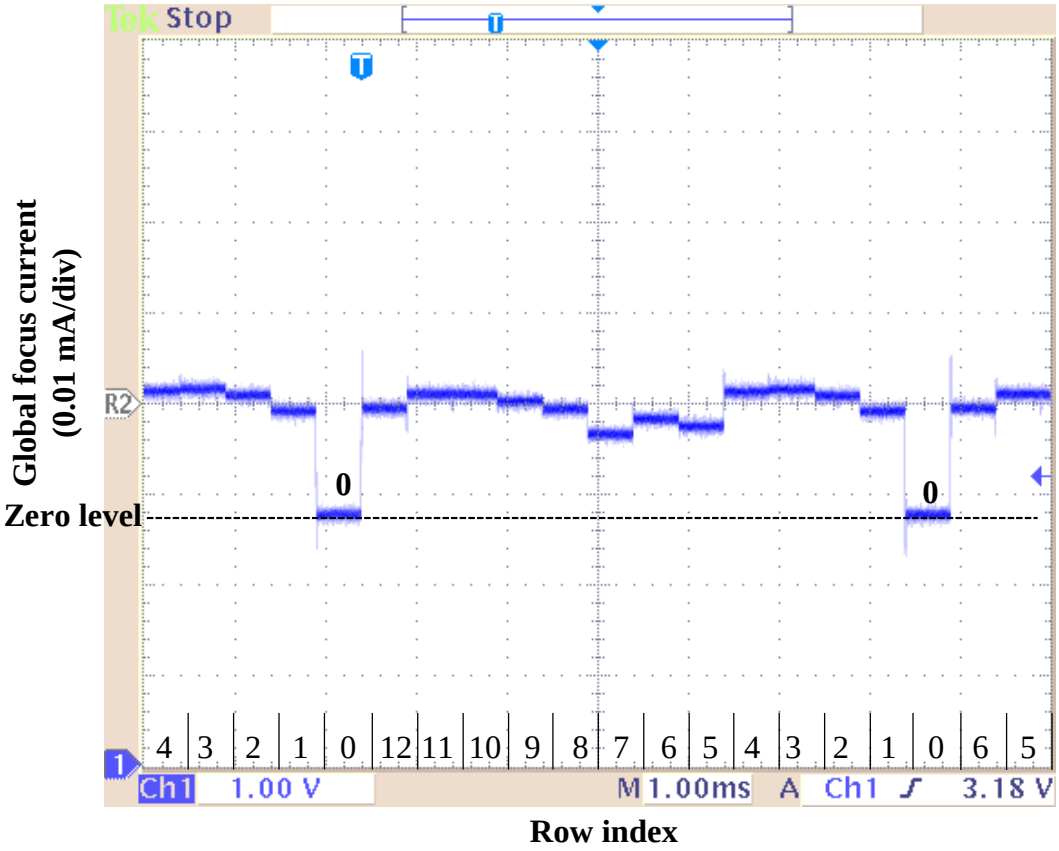


Fig. 68. Oscilloscope traces of global focus currents measured for all 12 rows when there is no contrast in the image.

Given in Fig. 68 are the focus measure currents for all 12 rows when there is no contrast on the captured image. Based on theory, the output currents of all rows should equal. The results shows that current offset due to the mismatches between rows is less than $\pm 1.5 \mu A$. As we have discussed before, if we have more cells per line, SNR will increase. From Fig. 68, we can see that our technique is based on the image depth and high frequency spatial signals, just like all other passive range finding techniques.

Fig. 69 shows the global focus currents measured for all 12 rows when testing board1 was put $420cm$ away with $f - stop = 2.8$. We can see that the focus measure difference between the winner and the closest lines shrinks. This also verifies the

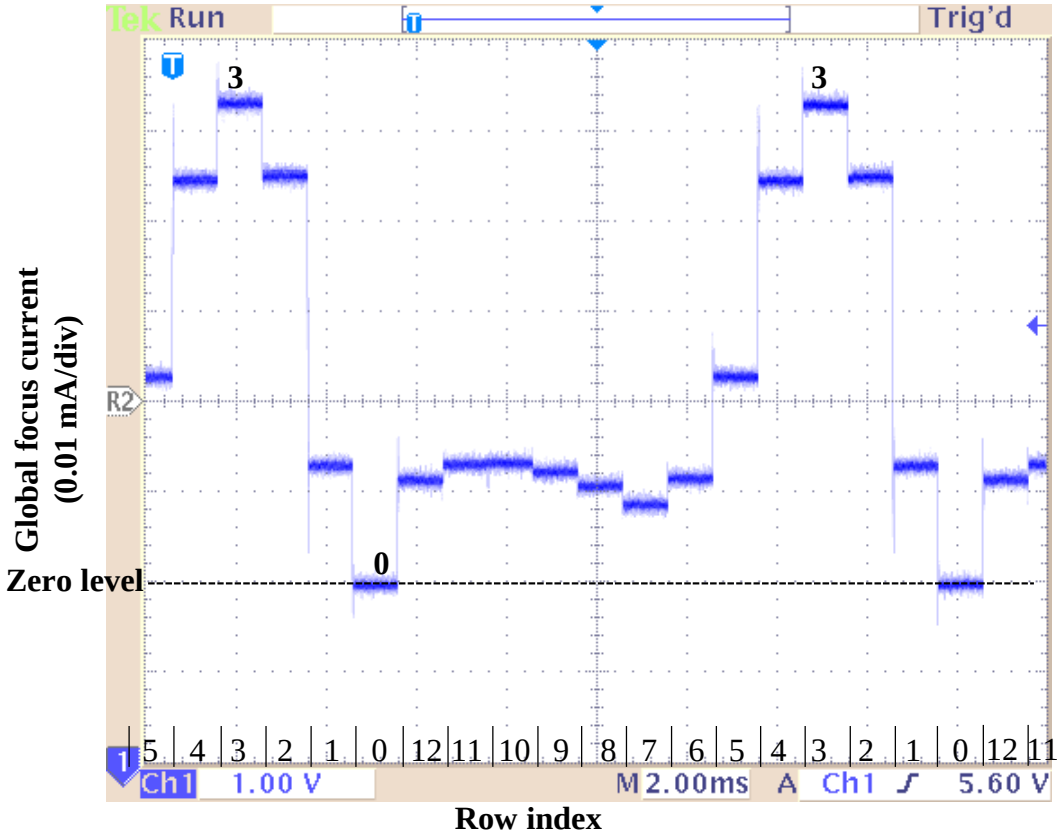


Fig. 69. Oscilloscope traces of global focus currents measured for all 12 rows when testing board1 was put 440cm away with $f - stop = 2.8$.

design optimization that was discussed in Chapter III.

Fig. 70 shows the global focus currents measured for all 12 rows when testing board2 was put 440cm away with $f - stop = 1.4$. Because the pattern contrast on the testing board2 is colorful and not as sharp as that on testing board1, we can see that the output current difference between the winner and the closest lines shrinks. This is due to two factors as discussed in Chapter III: (a) The sensor is insensitive to colorful pattern. (2) The spatial frequency of testing board2 is too high so that the lens filters some high-frequency edges.

Fig. 71 shows the global focus currents measured for all 12 rows when testing board1 was put 440cm away with $f - stop = 1.4$ and under low illumination. Com-

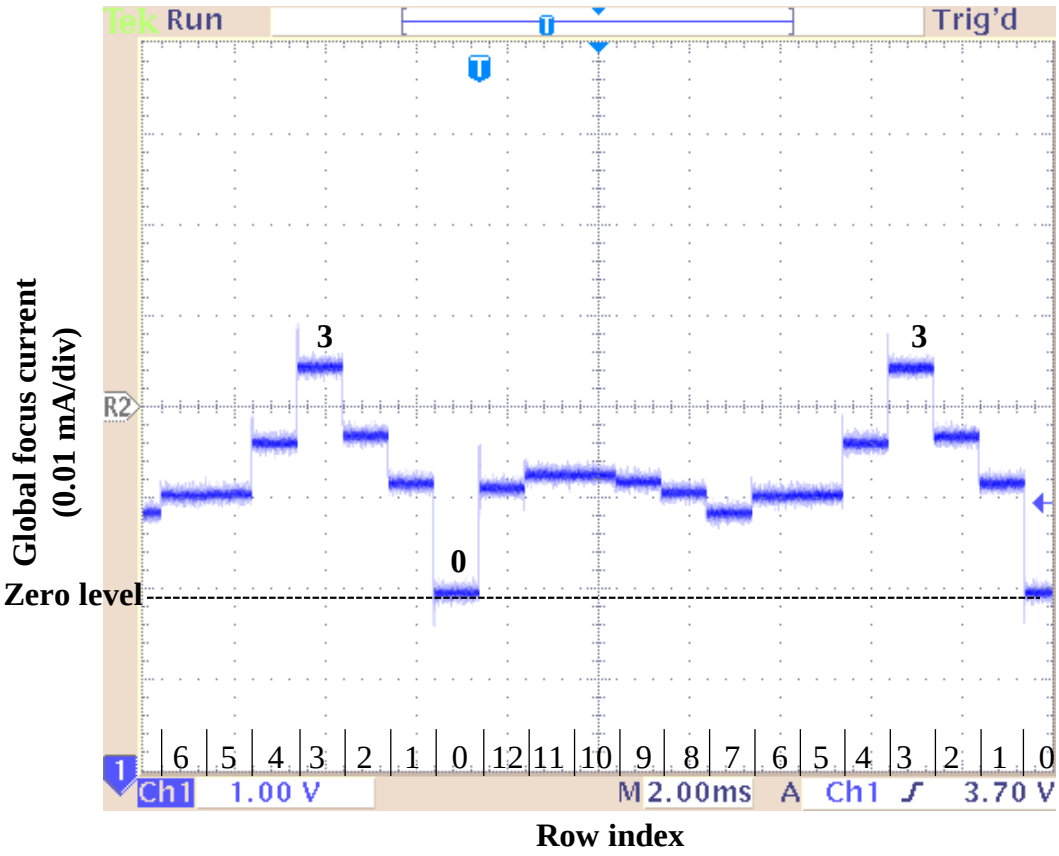


Fig. 70. Oscilloscope traces of global focus currents measured for all 12 rows when testing board2 was put $440cm$ away with $f - stop = 1.4$. Note that the prototype is insensitive to colorful pattern.

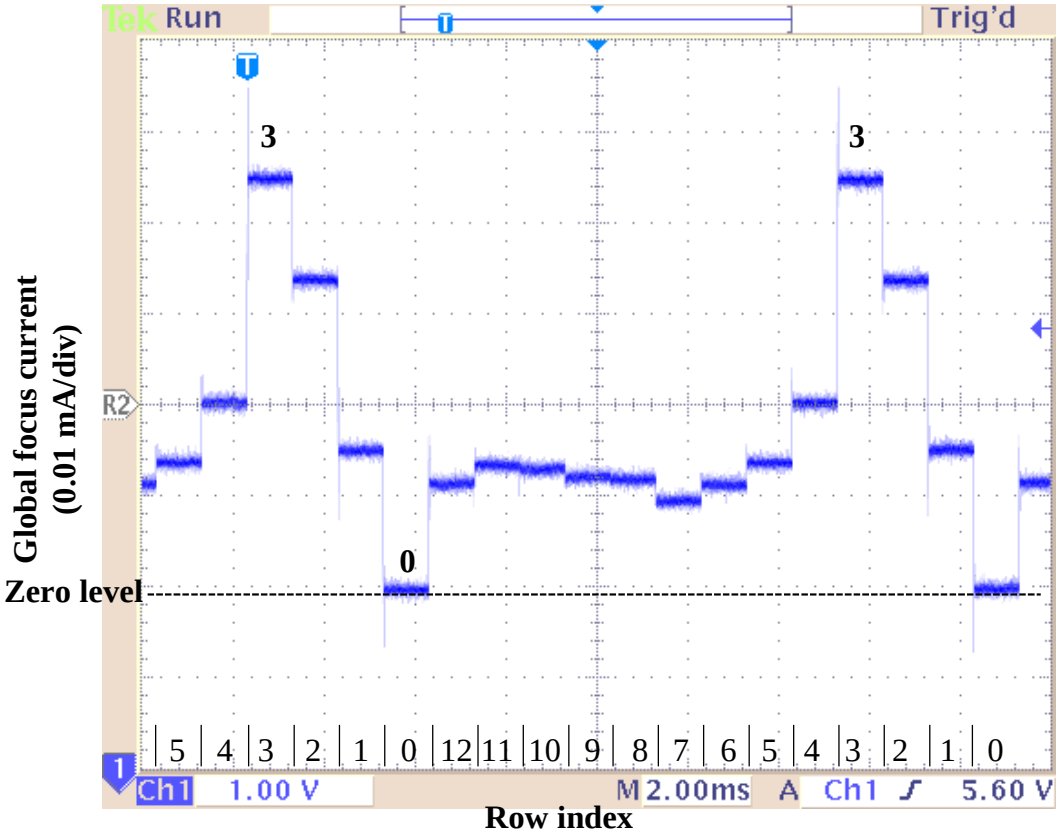


Fig. 71. Oscilloscope traces of global focus currents measured for all 12 rows when testing board2 was put 420cm away with $f - stop = 1.4$. The lumination is decreased by more than 50% compared with the environmental setup Fig. 67.

pared with the environmental setup Fig. 67, the lumination was decreased by more than 50%. We can see that the focus measure difference between the winner and the closest line doesn't shrink much. This verifies the performance of the logarithmic photosensors which only detect the illumination ratio between the neighboring pixels rather than the absolute illumination difference. Thus, this prototype is a very good quality that it's insensitive to environmental illumination variation.

When the object distance is between the nominal range sensed by the two neighboring rows, these two rows may have the same focus measure. As shown in Fig. 72, row9 and row10 have the same output current. This verifies the discussion in Chapter

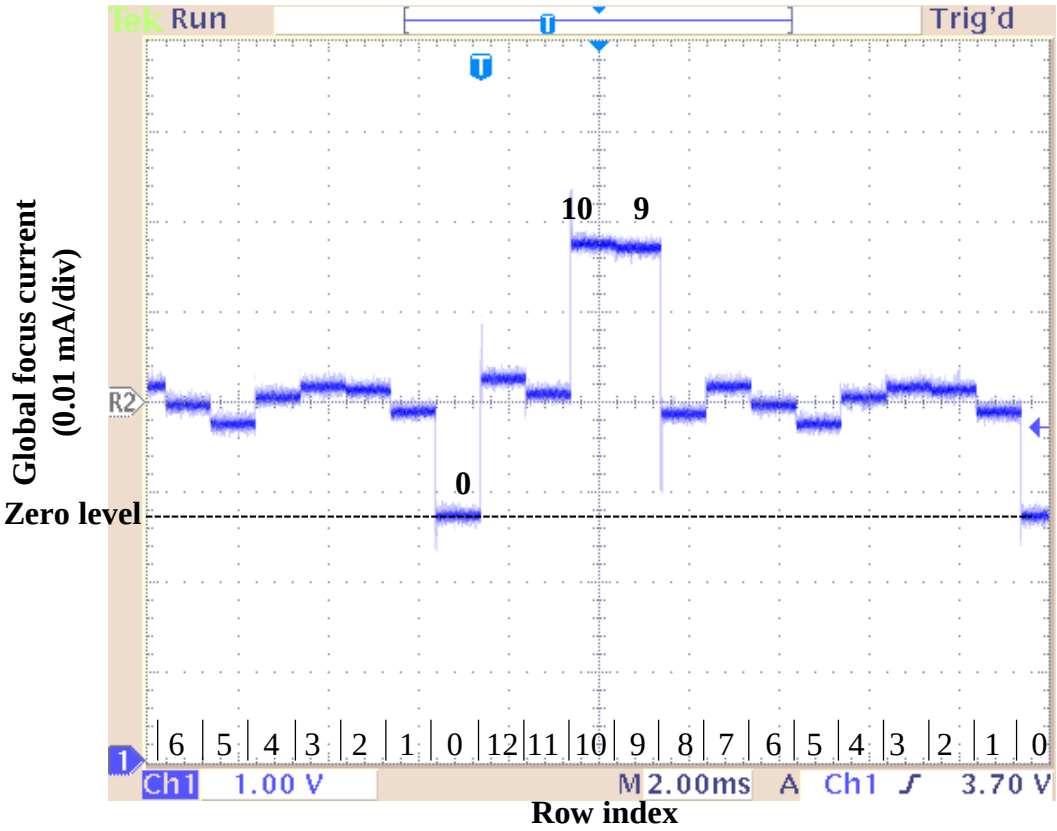


Fig. 72. Oscilloscope traces of global focus currents measured for all 12 rows when testing board2 was put 210cm away with $f - stop = 1.4$.

III. In this case, WTA may have two winners, which correctly shows the range of the object.

3. Result Summary

The experimental results are summarized in Table VIII. The results are consistence with the theory. Power dissipation and the active area of the sensor/processor chip are small.

Table VIII. Experimental results of the first prototype

Transmitter	Unnecessary
Mechanical adjustment	Unnecessary
A/D converter, memory, CPU	Unnecessary
Chip active area	$1.8mm \times 1.44mm$
Power supply and dissipation	$5V$ and $30mw$
Lens	Pentax $50mm/f1.4$ lens
Range	$1m$- $5m$
Speed	No visible delay. ($< 1ms$ based on
	Spice simulation under office illumination)

CHAPTER VI

CONCLUSION, LESSONS AND FUTURE WORK

In this dissertation a novel continuous-time vision-based range-finding technique is presented. The mathematical model has been built and verified using both Matlab simulator and hardware emulator. The proposed technique determines the range by identifying the location of best focus over the sensor plane without modulating any camera parameter. The competitive focus-searching process is free of confusion as long as the object occupies the entire field-of-view on the plane of non-orthogonal tilting and exhibits a spatially constant pattern. The signal processing tasks involved are simple, memoryless, and amenable to continuous-time implementation with a dedicated sensor/processor chip. This dissertation also gives the detailed circuit implementation of the chip.

All these efforts have culminated in a prototype system built with a custom Scheimpflug camera and a dedicated sensor/processor chip designed and fabricated in $0.5\mu m$ CMOS. The experimental results verified the expected functionality.

A. Lessons

What are the lessons from this system design? First, the system level design should start from a view of real-world and simple theories. *Keep it simple, keep it straightforward.* Gaussian lens law is the most fundamental law in almost all the vision-based range-finding technologies. Based on that, depth-from-focus methods are most direct methods and they can achieve much better resolution than depth-from-defocus methods that rely on specific optical models. The major constraint that sacrifices the cost and speed of DFF and DFD methods is that multiple frames have to be captured and stored for the further digital analysis. Capturing multiple frames needs sophisticated

mechanical adjustment, which is much slower than expensive than electronic calculation; storing frames adds the cost of memory and control circuits; digital processing needs another CPU and makes the single-chip range-finder scheme difficult to realize. Therefore, the best choice is to extract range information from a single frame using an analog processor.

Second, analog world is never as perfect as digital world. The range-finding chip would not work without well-conditioned input from the carefully designed photosensors and well-adjust camera body. One common error is that the designers usually start from perfectly mathematical models and ideal inputs without thinking about the nonideal factors of noise, mismatch, optical nonideality, etc. When designing an active range-finder system, designers can always improve the quality of input even after the system was built and make the life easier. However, in passive system, designers must consider the nonideality in the beginning of design and strive to generate the best possible inputs for the system. Therefore, overdesign is very important because some nonideal factors are almost impossible (very difficult) to be included in simulation. Only the final testing in real world can tell the performance of the system.

B. Future Work

The first testing prototype based on the proposed technique has been built and verified. The testing results prove that it is suitable for the applications in collision-avoidance. Since the first chip uses logarithmic photo sensors whose speed is slow during very dark lumination to capture the image, its performance is sacrificed in some environment. Future work to improve the speed of photo sensor will benefit certain applications. Minimizing the offset between pixels also helps to improve the system resolution.

1

REFERENCES

[1] S. Middelhoek and S. Audet, *Silicon Sensors*, London: Academic Press, 1989.

[2] "35 mm and medium format lenses," Research report, Photodo.com, (http://www.photodo.com/nav/prodindex.html), 2002.

[3] T. Delbruck, "Investigations of visual transduction and motion processing," Ph.D. dissertation, California Institute of Technology, Pasadena, California, 1993.

[4] W. D. Jones, "Keeping cars from crashing," *IEEE Spectrum*, pp. 40–45, September 2001.

[5] "Special investigation report on highway rear-end collision prevention technologies," Research report, National Transportation Safety Board, May 2001.

[6] R. Resendez, "Motor vehicle crashes - data analysis and ivi program emphasis," Research Report 12025, U.S. Department of Transportation, ITS Joint Program Office, November 1999.

[7] "Automotive collision avoidance systems (ACAS) program final report," Research report, U.S. Department of Transportation, National Highway Traffic Safety Administration, 1998.

[8] W. R. Garrot, M. A. Flick, and E. N. Marzae, "Hardware evaluation of heavy truck side and rear object detection systems," Research Report 951010.

[9] D. Nitzan, A. E. Brain, and R. O. Duda, "The measurement and use of registered reflectance and range data in scene analysis," in *Proceedings of IEEE*, vol. 65, pp. 206–220, 1976.

[10] D. M. Gavrila, "Sensor-based pedestrian protection," *IEEE Intelligent Systems*, vol. 16, pp. 77–81, 2001.

[11] Yen-Fi Liu, "A unified approach to image focus and defocus analysis," Ph.D. dissertation, State University of New York, Stony Brook, New York, 1998.

[12] K. Konolige, "Small vision systems: Hardware and implementation," in *8th International Symposium on Robotics Research*, 1997.

[13] A. P. Pentland, "A new sense for depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, pp. 523–531, July 1987.

[14] A. P. Pentland, "A simple, real-time range camera," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, June 1989, pp. 256–261.

[15] K. G ́o tz, "Course-control, metabolism and wing interference during ultralong tethered flight in drosophila melanogaster," *J. Exp. Biol.*, vol. 128, pp. 35–46, 1987.

[16] N. Giaquinto, M. Savino, and S. Taraglio, "A CNN-based passive optical range finder for real-time robotic applications," *IEEE Trans. on Instrumentation and Measurement*, vol. 51, pp. 314–319, April 2002.

[17] W. J. Ballantyne, *Distance*, CRC Press LLC Online, www.crc.com, 2000.

[18] H. Hua, Y. Wang, and D. Yan, "A low-cost dynamic range-finding device based on amplitude-modulated continuous ultrasonic wave," *IEEE Trans. on Instrumentation and Measurement*, vol. 51, pp. 362–367, April 2002.

[19] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Tran. Pattern Analysis and Machine Intelligence*, vol. 16, pp. 824–831, August 1994.

[20] S. K. Nayar, "Shape from focus system," in *Proceedings CVPR '92., 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, 1992, pp. 302–308.

[21] S. K. Nayar and Y. Nakagawa, "Shape from focus: an effective approach for rough surfaces," in *IEEE International Conference on Robotics and Automation*, Sacramento, California, 1990, pp. 218–225.

[22] S. K. Nayar and M. Noguchi, "Microscopic shape from focus using active illumination," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, Jerusalem, Israel, 1994, pp. 147–152.

[23] M. Subbarao and T. K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Tran. Pattern Analysis and Machine Intelligence*, vol. 20, pp. 864–870, August 1998.

[24] M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Tran. Pattern Analysis and Machine Intelligence*, vol. 17, pp. 266–274, March 1995.

[25] D. Ziou, "Passive depth from defocus using a spatial domain approach," in *6th International Conference on Computer Vision*, Bombay, India, 1998, pp. 799–804.

[26] Y. Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: how different really are they?" in *14th International Conference on Pattern Recognition*, Brisbane, Australia, 1998, pp. 1784–1786.

[27] S. K. Nayar, M. Watanabe, and M. Noguchi, "Real-time focus range sensor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 1186–1198, December 1996.

[28] T.L Hwang, J. J. Clark, and A. L. Yuille, "A depth recovery algorithm using defocus information," in *Proceedings CVPR '89., IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, 1989, pp. 476–482.

[29] T. Wei, "Three-dimensional machine vision using image defocus," Ph.D. dissertation, State University of New York, Stony Brook, New York, 1998.

[30] "Lux and light - illumination, exposure, and sensitivity," Research report, Micron, http://www.micron.com/imaging/Technology/Lux_and_Light.

[31] B. K. P. Horn, *Robot Vision*, Cambridge: The MIT Press, McGraw-Hill Book Company, 1987.

[32] A. El Gamal, *Classnote for EE392B*, Department of Electrical Engineering, Stanford University, 2001.

[33] "Best of eo application notes," Research report, Edmund Industrial Optics, vol. 1, pp. 10–25, 2002.

[34] V. Dragomirescu, O. Hockwin, H.R. Koch, and K. Sasaki, "Development of a new equipment for rotating slit image protography according to Scheimpflug's principle," *Interdiscipl. Top. Gerontol.*, vol. 13, pp. 118–130, 1978.

[35] S. Walker, "Two-axis Scheimpflug focusing for particle image velocimetry," in *Intl. Congress on Instrumentation in Aerospace Simulation Facilities*, Cleveland, Ohio, 2001, pp. 114–124.

[36] A. Krishnan and N. Ahuja., "Range estimation from focus using a non-frontal imaging camera," *Journal of Computer Vision*, vol. 20, pp. 169–185, 1996.

[37] "Depth of field for titled sensor surface," Research report, University of Illinois, Urbana, 1994.

[38] N. Koren, *Understanding image sharpness and MTF*, http://www.normankoren.com/index.html.

[39] W. Wetherell, "Afocal lenses," *Applied Optics and Optical Engineering*, vol. 10, pp. 110, 1987.

[40] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, New York: Wiley-Interscience, 1st edition, 1991.

[41] J. W. Goodman, *Introduction To Fourier Optics*, New York: McGraw-Hill Science/Engineering/Math, 2nd edition, 1996.

[42] G. R. Fowles, *Introduction To Modern Optics*, New York: Dover Pubns, 2nd edition, 1989.

[43] R. H. Vollmerhausen and R. G. Driggers, *Analysis of Sampled Imaging Systems*, Bellingham, Washington: SPIE Press, 2000.

[44] R. Kauert, W. Budde, and A. Kalz, "A monolithic field segment photo sensor system," *IEEE Journal of Solid-state Circuits*, vol. 30, pp. 807–811, July 1995.

[45] M. Ben-Chorin, J. Diener, D. Kovalev, G. Polisski, and F. Koch, "The temperature dependence of the absorption coefficient of porous silicon," *J. Appl. Phys.*, vol. 80, pp. 5978–5985, November 1996.

[46] E. R. Fossum, "CMOS image sensors: electronic camera-on-a-chip," *IEEE Trans. on Electron Devices*, vol. 44, pp. 1689–1698, October 1997.

[47] Dave Litwiller, "CCD vs. CMOS: Fact and fiction," *Photonics Spectra*, vol. 3, pp. 154–158, January 2001.

[48] S. K. Mendis, S. Kemeny, and E. Fossum, "CMOS active pixel image sensor," *IEEE Trans. on Electron Devices*, vol. 41, pp. 452–453, March 1994.

[49] G. P. Weckler, "Operation of p-n junction photodetectors in a photon flux integration mode," *IEEE J. Solid-State Circuits*, vol. sc-2, pp. 65–73, 1967.

[50] P. Noble, "Self-scanned solid image detector arrays," *IEEE Tran. Electron Devices*, vol. 15, pp. 202–209, April 1968.

[51] C. Wang, I. L. Fujimori, and C. G. Sodini, "A 256 × 256 CMOS differential passive pixel imager with fpn reduction techniques," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 2031–2037, December 2000.

[52] C. O. Staller R. H. Nixon, S. E. Kemeny and E. R. Fossum, "128 × 128 CMOS photodiode-type active pixel sensor with on-chip timing, control and signal chain electronics," in *Charge-Coupled Devices and Solid-State Optical Sensors V, Proc. SPIE*, San Jose, California, vol. 2415, pp. 117–123, 1995.

[53] S. Kavadias, B. Dierickx, D. Scheffer, A. Alaerts, D. Uwaerts, and J. Bogaerts, "A logarithmic response CMOS image sensor with on-chip calibration," *IEEE Journal of Solid-state Circuits*, vol. 35, pp. 1146–1152, August 2000.

[54] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, New York, Oxford: Oxford University Press, 4th edition, 1998.

[55] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, New York: John Wiley & Sons, Inc., 3rd edition, 1993.

[56] R. Sarpeshkar, J. Kramer, G. Indiveri, and C. Koch, "Analog VLSI architectures for motion processing: From fundamental limits to system applications," *Proceedings of the IEEE*, vol. 84, pp. 969–987, July 1996.

[57] R. Sarpeshkar, T. Delbruck, and A. A. Mead, "White noise in MOS transistors and resistors," *IEEE Circuits and Systems*, vol. 9, pp. 23–29, 1993.

[58] N. Tu, R. Hornsey, and S. G. Ingram, "CMOS active pixel image sensor with combined linear and logarithmic mode operation," in *IEEE Canadian Conference on Electrical and Computer Engineering*, Waterloo, Ontario, Canada, May 1998, vol. 2, pp. 754–757.

[59] S. E. Kemeny S. K. Mendis and E. R. Fossum, "A 128×128 CMOS active pixel image sensor for highly integrated imaging systems," in *International Technical Digest. Electron Devices Meeting*, Washington, DC., Dec 1993, pp. 583–586.

[60] U. Çilingiroğlu and S. Chen, "A CMOS imaging chip for real-time range finding," in *Proc. 43rd IEEE Midwest Symp. on Circuits and Systems*, Lansing, Michigan, 2000, vol. 3, pp. 1124–1127.

[61] U. Çilingiroğlu, S. Chen, E. Çilingiroğlu, and B. Siddik Yarman, "A single-chip vision-based range sensor," in *Proc. IASTED Intl. Conf. Signal Processing, Pattern Recognition and Application*, Crete, Greece, 2002, pp. 413–418.

[62] U. Çilingiroğlu, E. Çilingiroğlu, S. Chen, and B. Siddik Yarman, "Real-time range sensing with a Scheimpflug camera and a single custom sensor/processor chip," *accepted for presentation at 2002 /S&T/SPIE Symp. Electronic Imaging Science and Technology*, San Jose, California, January 2003.

[63] B. Razavi, *Design of Analog CMOS Integrated Circuits*, New York: McGraw-Hill Education, 2000.

[64] D. Johns and K. Martin, *Analog Integrated Circuit Design*, New York: John Wiley & Sons, Inc., 1997.

[65] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, and C. A. Mead, "Winner-take-all network of o(n) complexity," in *Advances in Neural Information Processing Systems I*, Morgan Kaufmann, San Mateo, California, 1989, pp. 703–711.

[66] B. Şekerkiran and U. Çilingiroğlu, "Improving the resolution of Lazzaro winner-take-all circuit," in *International Conf. on Neural Networks (ICNN97)*, Houston, 1997, vol. 2, pp. 1005–1008.

[67] M. E. Robinson, H. Yoneda, and E. Sanchez-Sinencio, "A modular CMOS design of a hamming network," *IEEE Trans.*, vol. 3, pp. 444–456, 1992.

[68] J. A. Starzyk and X. Fang, "CMOS current mode winner-take-all circuit with both excitatory and inhibitory feedback," *Electron Letter*, vol. 29, pp. 908–910, 1993.

[69] K. R. Laker and W. M. C. Sansen, *Design of Analog Integrated Circuits and Systems*, New York: McGraw-Hill, 1994.

APPENDIX A

DERIVATION OF PSF ON THE TILTED PLANE

In Chapter 3, we use the ellipse point spread function (PSF) to represent the blur distribution on the tilted plane. This appendix will give the detailed derivation of the PSF on a tilted plane.

First let's review the PSF on a vertical plane. Shown in Fig. 73 is the basic image formation geometry. Each point on the object plane is projected onto a single point in the image plane, thus causing a clear image $I_f(x, y)$. For example, point N, P on the object plane will cause clear points I, Q in the image plane. N is also on FOP that was discussed in Chapter 3. When plane V does not coincide with the image plane and is displaced from it by a distance δ, the energy received from each point in the object plane is distributed over a circular. It's well known that the distribution function $h_V(x, y)$ on the vertical plane V around blur center is very often described with [20]:

$$h_V(x, y) = \frac{1}{2\pi\sigma_h^2} e^{-\frac{x^2+y^2}{2\sigma_h^2}} \tag{A.1}$$

$$\sigma_h = kr = k\frac{\delta}{2N} \tag{A.2}$$

where k is a constant smaller than 1 and mostly decided by the lens quality, r is the patch radius, and δ is the plane displacement. The PSF is the same for all points on the object plane.

Now consider the image plane intersects a tilted plane T intersecting plane V at o as shown in Fig. 74 where the configuration is viewed along axis x. Since N is on FOP at the same time, it projects a perfectly focused image I as the intersection point of plane T and image plane. On the other hand, the flux received from the

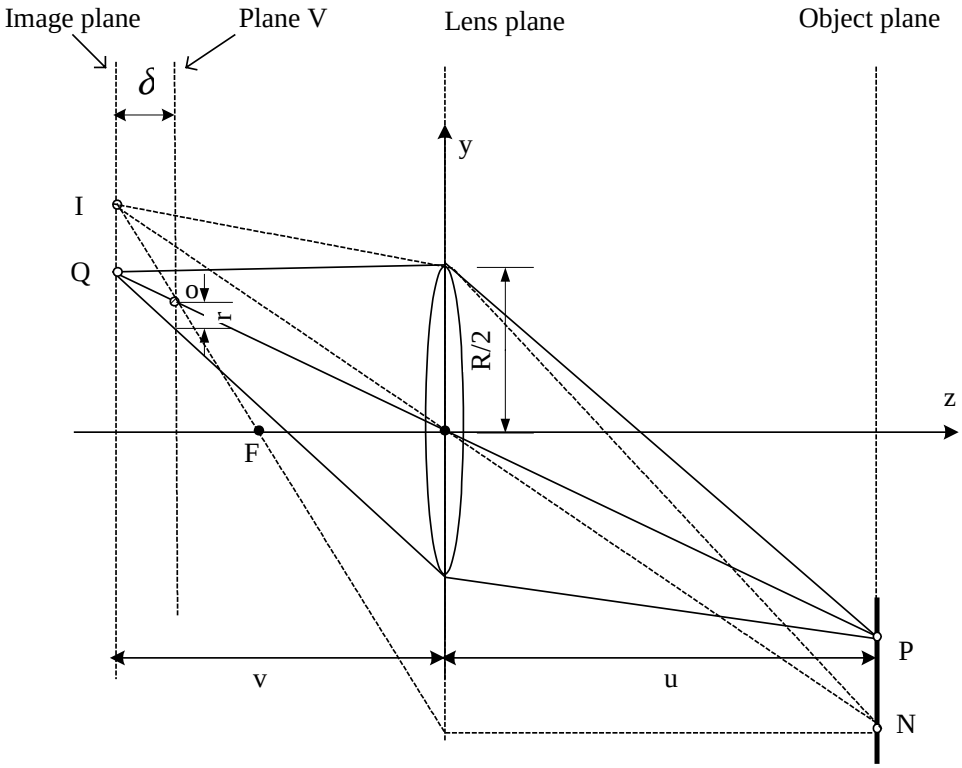


Fig. 73. Formation of defocused image on a vertical plane.

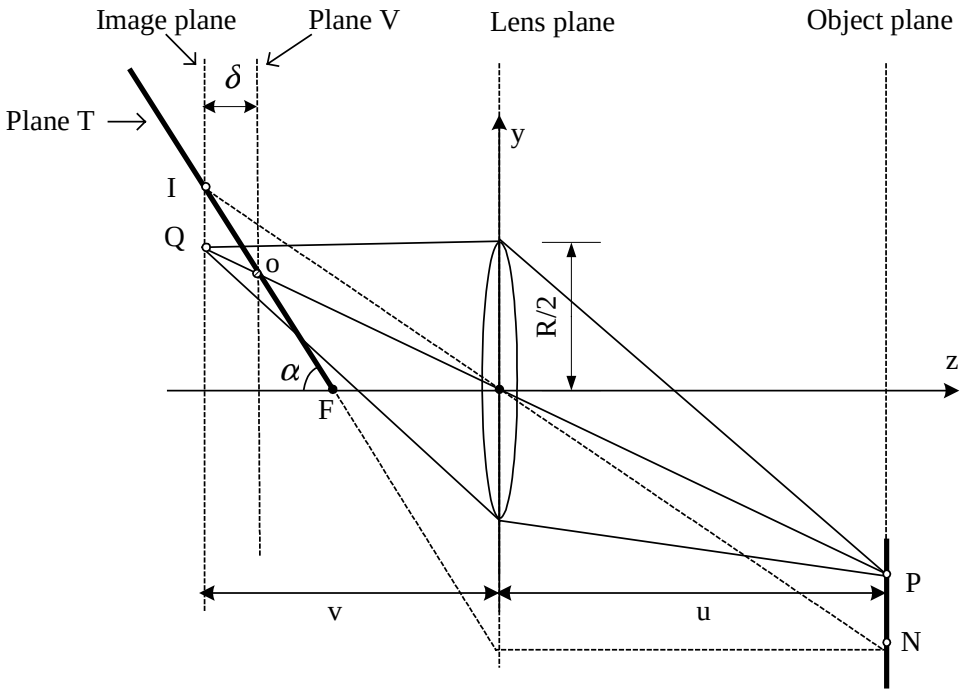


Fig. 74. Formation of defocused image on a tilted plane.

object point P by the lens is distributed over a blurred patch on the tilted plane T, just as what happens on the vertical plane V (see Fig. 73). But we can see that the PSF is not the same for all points on the object plane. For example, PSFs of point N and the point P are surely different. This is the foundation of our proposed technique to extract range information from just one image. A cone model is used to derive the PSF for point P. Since all the energy of P collected by the lens can be thought to be focused on point Q, we can assume all the flux are emitted by Q.

To simplify the PSF derivation on plane T, first let's assume that the coordinate of point Q is $(0, 0, Z_0)$. We will remove this constraint later on. Shown in Fig.75 is a 3D cone model combined by the lens and Q. Plane V and plane T intersect the cone. From now on, when we mention plane V, T, we means only the intersection part of plane V and T. We can easily found that energy collected by plane V equals to that collected by plane T and equals to that collected by the lens. $h_V(x_v, y_v)$ is

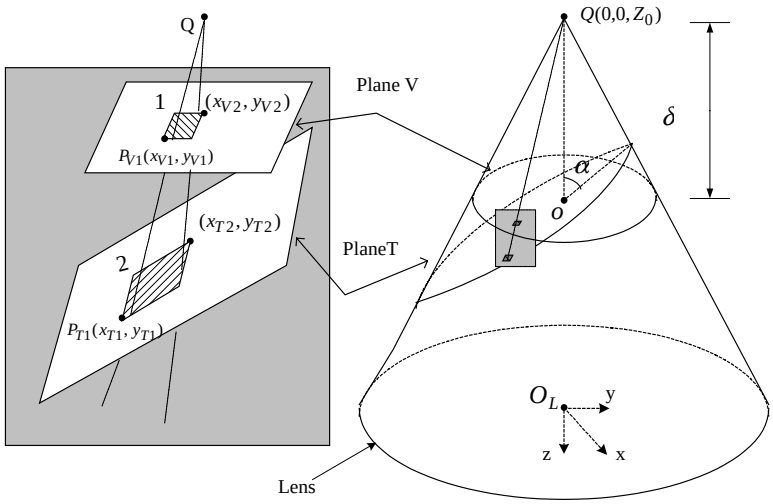


Fig. 75. The cone model corresponding to the 2D graph in Fig. 74. Q is the focus image point; plane V is a sensor plane parallel to the image plane with displacement δ; plane T is the tilted sensor plane that intersects plane V at node o. Pixel 2 is the projection of pixel 1 from plane V to plane T. O_L is the optical center of the lens.

the PSF of point P on the vertical plane V and $h_T(x_t, y_T)$ is the PSF on the tilted sensor plane T. (X, Y, Z) is the $3D$ coordinate, (x_T, y_T) is the $2D$ coordinate used in T plane and (x_V, y_V) is the 2D coordinate used in plane V. Line QO_L intersection plane V at point o, where O_L is the optical center of the lens.

Plane V can be represented by $z = Z_1$ where $|Z_0 - Z_1| = \delta$ and the coordinate of point o is $(0, 0, Z_0 + \delta)$. Pixel 1 is a small rectangular $(x_{V1}, y_{V1}, x_{V2}, y_{V2})$ on plane V; pixel 2 $(x_{T1}, y_{T1}, x_{T2}, y_{T2})$ on plane T is the projection of pixel 1. The enlarged graph of projection from pixel 1 to pixel 2 is shown in the upper-left side of Fig. 75.

Given the lens system, coordinate of Q, tilted sensor plane T, we can map the PSF in the tilted plane V to that in the vertical plane V.

As we can see in Fig. 75, the flux E_1 received by pixel 1 in plane V equals to the flux E_2 received by pixel 2 in plane T. If the area of the patch is small enough, we

have the following equations.

$$E_1 = h_T(x_{V1}, y_{V1})(x_{V2} - x_{V1}) = h_T(x_{V1}, y_{V1})dx_V dy_V \tag{A.3}$$

$$E_2 = h_T(x_{T1}, y_{T1})(x_{T2} - x_{T1}) = h_T(x_{T1}, y_{T1})dx_T dy_T \tag{A.4}$$

.

First let's find the mapping from a point $P_T(x_T, y_T)$ in plane T to P_V in plane V. The cooresponding $3D$ coordinates of P_T is $(x_T, y_T \cos\alpha, Z_1 - y_T \sin\alpha)$. We can see that line QP_T intersects plane V at P_V. Line QP_T can be represented by

$$\frac{X - x_T}{0 - x_T} = \frac{Y - y_T \cos\alpha}{0 - y_T \cos\alpha} = \frac{Z - (Z_1 - y_T \sin\alpha)}{Z_0 - (Z_1 - y_T \sin\alpha)} \tag{A.5}$$

Solve Equ. A.5, we can get the coordinate of $P_V(X_V, Y_V, Z_V)$:

$$X_V = \frac{\delta x_T}{\delta - y_T \sin\alpha} \tag{A.6}$$

$$Y_V = \frac{\delta y_T \cos\alpha}{\delta - y_T \sin\alpha} \tag{A.7}$$

$$Z_V = Z_1 \tag{A.8}$$

Since $x_V = X_V, y_V = Y_V$, we have the distribution on the plane T as:

$$h_T(x_T, y_T) = \frac{1}{2\pi\sigma_h^2} e^{-\frac{(\frac{\delta x_T}{\delta - y_T \sin\alpha})^2 + (\frac{\delta y_T \cos\alpha}{\delta - y_T \sin\alpha})^2}{2\sigma_h^2}} \frac{dx_V dy_V}{dx_T dy_T} \tag{A.9}$$

In a general cone model, the coordinate of image point Q is given by (X_Q, Y_Q, Z_Q). Since o is the intersection of line QO_L and plane V and $\delta << Z_0$ (δ is in the range of hundreds of μm and Z_0 is in the range of a few cm), the coordinate of $o(X_o, Y_o, Z_o)$ is given by

$$X_o = X_Q \cdot \frac{Z_0 - \delta}{Z_0} \approx X_Q \tag{A.10}$$

$$Y_o = Y_Q \cdot \frac{Z_0 - \delta}{Z_0} \approx Y_Q \tag{A.11}$$

$$Z_o = Z_0 - \delta \tag{A.12}$$

Follow the same procedure, we will get Equ. A.9. So Equ. A.9 is still a very good approximation for PSF on a tilted plane.

The projection from a infinitely small rectangular $Pixel1(x_{V1}, y_{V1}, x_{V2}, y_{V2})$ in plane T to $Pixel2$ in plane V is given by:

$$x_V = \frac{\delta}{\delta - y_T \sin\alpha} x_T, \Delta x_V = x_{V2} - x_{V1} \approx \frac{\delta}{\delta - y_{T1} \sin\alpha} \Delta x_T \tag{A.13}$$

$$y_V = \frac{\delta \cos\alpha}{\delta - y_T \sin\alpha} y_T, \Delta y_V = y_{V2} - y_{V1} \approx \frac{\delta \cos\alpha}{\delta - y_{T1} \sin\alpha} \Delta y_T \tag{A.14}$$

$$dx_V dy_V = \frac{\delta^2 \cos\alpha}{(\delta - y_T \sin\alpha)^2} dx_T dy_T \tag{A.15}$$

With Eqn. A.15 and Equ. A.9, we get $h_T(x_T, y_T)$ in the tilted plane T:

$$h_T(x_T, y_T) = \frac{1}{2\pi\sigma_h^2} e^{-\frac{\left(\frac{\delta y_T \cos\alpha}{\delta - y_T \sin\alpha}\right)^2 + \left(\frac{\delta x_T}{\delta - y_T \sin\alpha}\right)^2}{2\sigma_h^2}} \frac{\delta^2 \cos\alpha}{\delta - y_T \sin\alpha} \tag{A.16}$$

From Equ. A.2, we can see that most flux is focused in the area $|x, y| < \sigma_h = \frac{k\delta}{2N}$. We have

$$\frac{\delta}{\delta - y_T \sin\alpha} \approx 1 \quad \text{for} \quad |Y_T| < \sigma_h \tag{A.17}$$

According to Equ. A.17, Equ. A.16 can be simplified as

$$h_T(x_T, y_T) = \frac{\cos\alpha}{2\pi\sigma_h^2} e^{[(y_T \cos\alpha)^2 + (x_T)^2]/(2\sigma_h^2)} \tag{A.18}$$

Using Fourier transform and Equ. A.18, we get the frequency domain image distribution of point Q:

$$H_T(\lambda_{y_T}, \lambda_{x_T}) = \cos\alpha e^{-\frac{\sigma_h^2}{2}[{\lambda_{x_T}}^2 + \left(\frac{\lambda_{y_T}}{\cos\alpha}\right)^2]} \tag{A.19}$$

where λ_{x_T} and λ_{y_T} is the spatial frequency in x_T and y_T axises, respectively.

When $\delta << Z_0$, $X_o \approx X_Q, \delta \approx QI \cdot \tan\alpha$. We have

$$\sigma_h = k \cdot \delta \frac{\delta}{2N} = \frac{k}{2N}(oI \cdot \sin\alpha) = \frac{k}{2N}(QI \cdot \tan\alpha) \tag{A.20}$$

$$\begin{aligned} H_T(\lambda_{x_T}, \lambda_{y_T}) &= \cos\alpha \exp\{-\frac{(kQI\tan\alpha)^2}{2(2N)^2}[\left(\frac{\lambda_{y_T}}{\cos\alpha}\right)^2 + \lambda_{x_T}^2\} & \text{(A.21)} \\ &= \cos\alpha \exp\{-\frac{(koI\cos\alpha)^2}{2(2N)^2}[\left(\frac{\lambda_{y_T}}{\cos\alpha}\right)^2 + \lambda_{x_T}^2\} & \text{(A.22)} \end{aligned}$$

where oI and QI is the distance from o and Q to I, respectively.

From Equ. A.22, we can see that for a ideal thin convex lens: (1) With a fixed $oI \neq 0$, when spatial frequency $\lambda_{x_T}, \lambda_{y_T}$ increases, $H_T(\lambda_{x_T}, \lambda_{y_T})$ drops exponentially. So the imaging system is a low pass filter. (2) With a fixed spatial frequency $\lambda_{x_T}, \lambda_{y_T}$, when oI (the distance between the defocused center and focused point in plane T) increases, $H_T(\lambda_{x_T}, \lambda_{y_T})$ drops exponentially. So the the cut-off frequency of imaging system decreases sharply as the distance between the focused line and defocused line increases.

APPENDIX B

MATLAB CODES

```
%XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
% generate the image on a tilted plane
%XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
clear;
%The unit used in this program is 1um.
%Compared with tilt1-4.m, this program uses a simplified mathematical model.
%The size of each pixel is 20*20 um.
%The sampling step is 5 by 5um, for simplicity,
%we assume that in this 5*5 um square, the distribution is even .
%The simulated chip size is 2mm*2mm.

    stepsize=5; theta=pi/4; imagesize=200;
nx=-imagesize:stepsize:imagesize;
m1=nx';
[x1, y1] = meshgrid(nx, nx); [u,v]=size(nx);
u=v;

    %The object pattern is a sine wave.
z=(1+sin(x1.*0.04*pi));
figure(1);
clf;
mesh(x1,y1,z');
```

```
chipsize=round(imagesize/cos(theta)/stepsize)*stepsize;

mx=-chipsize:stepsize:chipsize;
my=mx';
[x1, y1] = meshgrid(mx, my); [Iu,Iv]=size(mx);
sen(Iv,Iv)=0;

%center of the image plane in matrix sen is (iu0,iv0)
iu0=(u+1)/2; iv0=(v+1)/2;
%center of the sensor plane in matrix sen is (su0,sv0)
su0=(Iv+1)/2; sv0=(Iv+1)/2;
%The mapping from image plane to sensor plane is:
%(j-iv0)=(j_sen-sv0); j_sen=j-iv0+sv0;
%(i-iu0)=(i_sen-su0); i_sen=round((i-iu0)/cos(theta))+su0;

i_sen=i-iu0+su0;
for i=1:u
matchi=cal_sim(i,nx,stepsize,theta);
for j=1:v
match= z(i,j)*matchi; j_sen=j-iv0+sv0;
i_sen=round((i-iu0)/cos(theta))+su0;

sen=addmatrix_sim(sen,i_sen,j_sen,match);

end
end
```

```
sen=sen';
figure(2);
clf;
mesh(x1,y1,sen);

%XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
% function cal
%XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

function match=cal(i,nx,stepsize,theta)
%the unit is um
%this function is to find the mapping matrix
%from a focused image point P(x0,y0) to the sensor plane.
%The original point is to be the intersection of
%image plane and sensor plane with y=0.
%delta is the Z-distance from P to the sensor plane.
%delta=x*tan(theta)
%For the same x, delta is same, so the matching matrix is same.
%for different x, the sizes of matching matrixes are different.

z0=5.05e4; f=5e4; R=2e4;
x=nx(i);
delta=abs(x*tan(theta))
r=delta*R/z0;
if(r < stepsize/2)match(1,1) = 1;
else
```

$sigma = r/(2^{0.5})$;

%calculate the size of the blurring matrix

%the area to be calculated is limited to be 3 sigma1.

%the size of matrix is choose to be 3simga1/stepsize.

$q = 2 * round(3 * sigma/stepsize/cos(theta)) + 1$;

match(q,q)=0;

%(Ix0, Iy0) is the center of the matching matrix.

Ix0=round(3*sigma/stepsize)+1; Iy0=Ix0;

for i=1:q

for j=1:q

x=stepsize*(i-Ix0); y=stepsize*(j-Iy0);

$Ix = delta * cos(theta) * x/(delta - x * sin(theta))$;

$Iy = delta * y/(delta - x * sin(theta))$;

%match(i,j) is the probability of illumination in the square pixel(i,j)

%Note that the mapping from tilted plane to vertical plane changes the area.

$pixelsize = stepsize^2 * (delta/(delta - x * sin(theta)))^2 * cos(theta)$;

$match(i, j) = pixelsize * 1/(2 * pi * sigma^2) * exp(-(Ix^2 + Iy^2)/(2 * sigma^2))$;

end

end

end %endof if

%XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

% function addmatrix_sim

%XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

```
%do convolution
%Add matrix "match" to matrix "sen"
function sen=addmatrix(sen,i,j,match)
[u, v] = size(sen);
[mu, mv] = size(match);
cu=round((mu+1)/2);
cv=round((mv+1)/2);
if(i¿cu-1) l_lim=cu-1; else l_lim=i-1; end;
if(j¿cv-1) t_lim=cv-1; else t_lim=j-1; end;
if(i¡=u-cu+1) r_lim=cu-1; else r_lim=u-i; end;
if(j¡=v-cv+1) b_lim=cv-1; else b_lim=v-j; end;
sen(i-l_lim:i+r_lim,j-t_lim:j+b_lim)=sen(i-l_lim:i+r_lim,j-t_lim:j+b_lim) ...
+match(cu-l_lim:cu+r_lim,cv-t_lim:cv+b_lim);

%XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
% Generate defocused image on a nontilted plane
%XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

% Generate defocused image on a vertical sensor plane
clear;
z0=5.05e4;f=5e4;R=2e4;theta=-pi/4;
u=5e3; delta=140;
r=delta*R/z0;
sigma = r/2^0.5;

stepsize=5; theta=pi/4; imagesize=200;
```

nx=-imagesize:stepsize:imagesize;

m1=nx';

$[x1, y1] = meshgrid(nx, nx)$; $[u, v] = size(nx)$;

u=v;

%The pattern is a sine wave.

z=(1+sin(x1.*0.02*pi));

figure(1);

clf;

mesh(x1,y1,z);

mapsize=round(2*r/stepsize)*stepsize;

$pixelsize = stepsize^2$; h(mapsize/stepsize*2+1,mapsize/stepsize*2+1)=0;

for x=-mapsize:stepsize:mapsize

i=(x+mapsize)/stepsize+1;

for y=-mapsize:stepsize:mapsize

j=(y+mapsize)/stepsize+1;

$h(i, j) = pixelsize * 1/(2 * pi * sigma^2) * exp(-(x^2 + y^2)/(2 * sigma^2))$;

%Here we can use 2D convolution

end

end

figure(2);

clf;

n2=-mapsize:stepsize:mapsize;

$[x2, y2] = meshgrid(n2, n2)$;

```
mesh(x2,y2,h);

chipsize=mapsize+imagesize;
mx=-chipsize:stepsize:chipsize;
my=mx';
[x1, y1] = meshgrid(mx, my); [Iu, Iv] = size(mx);
sen(Iv,Iv)=0;

%center of the image plane in matrix sen is (iu0,iv0)
iu0=(u+1)/2; iv0=(v+1)/2;
%center of the sensor plane in matrix sen is (su0,sv0)
su0=(Iv+1)/2; sv0=(Iv+1)/2;
for i=1:u
for j=1:v
match= z(i,j)*h;
j_sen=j-iv0+sv0;
i_sen=i-iu0+su0;
sen=addmatrix_sim(sen,i_sen,j_sen,match);
end
end

n3=-(imagesize+mapsize):stepsize:(imagesize+mapsize);
[x3, y3] = meshgrid(n3, n3);
figure(3);
clf;
mesh(x3,y3,sen);
```

```
%XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
% Photo processing
%XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
image(pic);
figure(2);
clf;
/hold on;
/line1=80; line2=170;
plot(pic(line1,:,1));
%f=8, find the focused line index
fp=6; f=fp*1000/36/12; period=4*12/fp
head=210; tail=round(head+1.8*period);

MTF0(line2-line1+1)=0;
for line=line1:line2
a=min(pic(line,head:tail,1));
b=max(pic(line,head:tail,1));
MTF0(line-line1+1)=(double(b)-double(a))/255;
end
```

$[MT, f\_index] = max(MTF0(1 : line2 - line1))$

VITA

Sicheng Chen was born in Fuzhou, Fujian province, China on April 13, 1974. In 1991, he entered Huazhong University of Science and Technology. In 1995, Sicheng received a Bachelor of Engineering in computer engineering and graduated with honors from department of computer science.

In September 1995, Sicheng entered the Institute of Computing Technology, Chinese Academy of Sciences in Beijing, China. His research area was focused on digital circuit testing and faulting simulation. In July of 1995, he obtained the Master of Science degree with his thesis "Fault simulation in synchronized sequential circuit". In September 1998, Sicheng entered the electrical engineering department at Texas A&M University to study analog and mixed-mode VLSI design under the advisement of Dr. Uğur Çilingiroğlu and Dr. Edgar Sanchez-Sinencio. Here he designed and fabricated several integrated circuits for range finding. He was a teaching assistant for several classes including filter design, electronics and electronic circuits. From May of 2001 to December of 2001, he was a co-op student in the wireless department of Texas Instruments Inc. He helped to design low-dropout regulators and other power management circuits. Sicheng's major field of interest and study continues to be analog and mixed-mode IC design. Presently he is working on adaptive systems and power management.

The typist for this thesis was Sicheng Chen.