PEOPLE'S DIMOCRATIC REPUBLIC OF ALGERIA

Ministry of Higher Education and Scientific Research

UNIVERSITY KASDI MERBAH OUARGLA

Faculty of New Information Technologies and Communication

Department of Electronics and Telecommunications

MASTER ACADEMIC MEMORY

Domain: Electronics and telecommunications

Field: Electronic

Specialty: Embedded Systems

Presented by:

AHMED MESSOUD ASMAA **TEMMAR RADJA**

Theme

Face Detection by the use of a Plenoptic Camera

Publicly supported

IN: 25/06/2018

Discussed front of the jury composed of:

Noun	*Grade*	*Quality*	*University*
W. BENZAOUI	**MAA**	**Supervisor**	**UKM OUARGLA**
Gh. ZIDANI	**MCB**	**A. Supervisor**	**UB2 BATNA**
F. CHARIF	**MCB**	**President**	**UKM OUARGLA**
Dj. SAMAI	**MCA**	**Examiner**	**UKM OUARGLA**

Academic year: 2017/2018

Dedications

I dedicate this modest memory:

To my very dear parents: my very dear mother who has supported me by these blessings and my dear father who has also been a moral and material help.

To my dear sisters "Ratiba" and "Ghania"

To my dear brothers "Bilal" and "Ahmed"

To all my family, big and small, whoever they are

To my sincere friends" Samiha" and "Radja"

To all my friends in the class of Embedded Systems

Ahmed Messaoud Asmaa

Dedications

I would to thank my parents who were and still always supporting me , Mom and dad Thank you for pushing me towards my goals , Thanks for you being by my side every time, I feel so honored and blessed to have you as my parents , and want to express my gratitude for your care and support over the years

To my two lovely sisters "Bouthaina" and "Safa"

To all my family

My sincere friend "Asmaa"and "khadidja" ,to all my friends

& to my friends in class of "Embedded Systems"

To all my teachers from the primary to the university to everyone who's Taught Me something, thank you

Temmar Radja

Acknowledgements

First and foremost, We would like to fully acknowledge and thank Allah for all the blessings and guidance we are granted, we thank Allah from the deepest of our hearts for enlightening us on the right path..

We want to thank our families for their emotional contributions and sacrifices. We would like to thank our supervisor who helped us to carry out this modest work in better conditions: Mrs. BENEZZAOU WAFA and for her interest and support, her great availability and her many advices during the writing of our thesis. We thank the Jury members Mr. F. CHARIF and Mr. A.SAMAI, who do us the honor of accepting to judge our work. Without forgetting of course all the teachers who contributed by their knowledge and their encouragement along our courses to everyone who's helped us in this work.

Résumé

L'imagerie de champ lumineux est apparue comme une technologie permettant de capturer des informations visuelles plus riches de notre monde. Contrairement aux caméras conventionnelles, qui enregistrent la projection 2D des rayons lumineux en intégrant angulairement les rayons à chaque pixel, les caméras à champ lumineux capturent non seulement l'intensité lumineuse dans le plan image mais aussi l'information sur la direction des rayons lumineux. Bien que la détection de visage soit une forme largement acceptée de biométrie en raison de sa facilité de capture, elle souffre toujours d'une imagerie floue en raison de la profondeur de champ limitée des systèmes d'imagerie 2D conventionnels. Dans ce travail, et afin de surmonter un tel défi, nous utilisons le champ lumineux pour obtenir un certain nombre d'images de profondeur. Ensuite, une technique de fusion est utilisée pour fusionner les meilleures images de mise au point. L'image obtenue contient toutes les faces de différents foyers présentés dans la même image, ce qui résout le problème de la détection de visage à partir de l'imagerie non focalisée.

Abstract

Light field imaging has emerged as a technology allowing to capture richer visual information from our world. Unlike conventional cameras, which record the 2D projection of the light rays by angularly integrating the rays at each pixel, a light field cameras capture not only the light intensity in the image plane but also the information about the direction of light rays. Although face detection is a widely accepted form of biometrics due to its ease of capture, it still suffers from out-of-focus imaging due to the limited depth-of-field found in conventional 2D imaging systems. In this work, and in order to overcome such a challenge, we use the light field to obtain a number of depth images. Then, a fusion technique is used to fuse the best focus images. The resulted image contains all faces from different focuses presented in the same image, which solves the problem of face detection from out-of-focus imaging.

ملخص

لقد ظهر التصوير في مجال الضوء كإحدى التقنيات التي تتيح التقاط معلومات بصرية أكثر ثراءً من عالمنا. على عكس الكاميرات التقليدية ، التي تسجل إسقاط بعدين فقط من أشعة الضوء عن طريق دمج أشعة الشمس في كل بكسل ، تلتقط الكاميرات الحقلية الضوء ليس فقط كثافة الضوء في مستوي الصورة ولكن أيضا معلومات حول اتجاه الأشعة الضوئية. على الرغم من أن اكتشاف الوجه هو شكل مقبول على نطاق واسع من القياسات الحيوية نظرًا لسهولة التقاطه ، فإنه لا يزال يعاني من التصوير خارج التركيز نظرًا لحدود العمق المحدودة الموجودة في أنظمة التصوير التقليدية ثنائية الأبعاد. في هذا العمل ، وللتغلب على هذا التحدي ، نستخدم حقل الضوء للحصول على عدد من صور العمق. ثم ، يتم استخدام تقنية الانصهار لدمج أفضل صور التركيز. تحتوي الصورة الناتجة على جميع الوجوه من بؤر مختلفة معروضة في الصورة نفسها ، والتي تحل مشكلة اكتشاف الوجوه من التصوير خارج التركيز

List of figures

Chapter 1

Light Field Imaging

Chapter 2

Images from the light field and digital refocusing

Chapter 3

Face detection and image fusion

Chapter 4

Results and Discussion

List of tables

Chapter 1

Light Field Imaging

Chapter 4

Results and Discussion

Contents

Chapitre 1

Light Field Imaging

Chapitre 2

Images from the light field and digital refocusing

Chapter 3

Face detection and image fusion

Chapter 4

Results and discussion

General introduction

Face detection presents a challenging problem in the field of image analysis and computer vision, and as such has received a great deal of attention over the last few years because of its many applications in various domains as in large scale surveillance applications.

In spite of many face capture devices (or cameras) and algorithms that are currently available, there exists still a wide spectrum of challenges (for example; pose, distance, presence of multiple faces, occlusions, etc.) that need to be addressed to achieve an accurate face biometric system. Among these challenges, identifying multiple faces with varying pose at various distance from the camera is still challenging. [RAG13]

In this work, we made an attempt to address the problem of detecting multiple faces various distance from the camera. However, the use of a conventional 2D face capturing camera exhibits the limitation of fixed point of view and a focus which is determined entirely on how the lens was set during the capture process. This will certainly limit the conventional camera when capturing the multiple 2D face images that are present in a given scene at various distances. One of the interesting ways to address this situation is by employing the light field camera.

Light field (or plenoptic) imaging is a technique which allows the capture of the 3-dimensional information in a scene using a single camera in a single exposure. This is done by adding an array of microlenses to a conventional camera such that it allows information about the angles of light rays as well as their intensity and position to be recorded. This information can be computationally post-processed to allow digital refocusing or camera repositioning. Such cameras have recently become commercially available and are mainly used in consumer photographic applications.

The employing of the light field camera permits us to obtain a set of images that can be used to perform refocus, all-in-focus image, depth estimation, synthetic aperture. We will try to acquire a set of light field images with different depths; than we will calculate the entropy of images to select the best two images to fuse. We will use three method of fusion DCT, Pyramid, Combination, the one gives as the best result we will work by it. And finally we will detect all the faces in the image by face detection algorithm in matlab.

Chapter 1

Light field imaging

1. Introduction

We present the technology of light field (LF) photography that's represents one of the new generation of imaging methods enabling in many ways rich features allowing more realistic, interactive, and immersive user experience. The basic idea of capturing light field data originates from the concept of integral photography in the beginning of twentieth century. Light field image acquisition process is based on capturing not only the light intensity in the image plane but also the information about the direction of light rays. These advanced features of light field imaging are likely to be exploited in many branches of research and industry, such as microscopy, computer vision, medicine, etc, in near future.

In this chapter we present the conventional imaging as well as problems of conventional imaging and what is the light field. Then, we represent the parameterization and plenoptic function of light field and the acquisition of the light field. We take the plenoptic camera Lytro use example and application of this camera.

2. Conventional imaging

Traditional cameras do not record most information about light distribution. Where the traditional imaging system is formed from a sensor image to optical parts that allow taking a picture. So that the image contains only one large lens that works a great deal on the capture of the optical beam (light rays) of the object and dropped on the sensors. The image had just one lens this lens sensor take all the rayon's light and we collect them.

When the object out of focus its image is blurred, and when the object on the focus plane its image is clear.

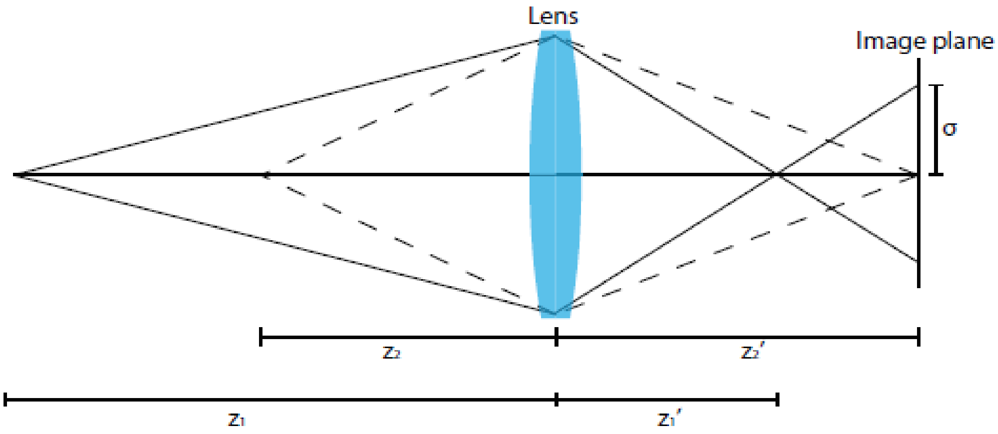


Figure 1.1. Defocused first order system.

Where z is the distance from the object to the front principal plane, z' is the distance from the rear principal plane to the image plane, and f is the focal length of the imaging system (1/f is commonly denoted as the power of the optical system).

2.1.Problems of conventional imaging

When we take an image by a conventional camera the object that is in the focal plane it will look pure and clear but the object that is out of the focal plane it will look unclear after the acquisition we can't take back any information about the out focal object this problem mean that the lens take just the rays that in the lens, dissociation of ray directions, this information is important to recover the image reconstruction and depth, so the conventional imaging do not record most of the information about the light entering the camera.

For example, if we think about the light deposited at one pixel, a photograph tells us nothing about how the light coming from one part of the lens differs from the light coming from another. It turns out that these differences are the crucial pieces of missing information that lead to the focus problem in conventional photography.

Figure 1.2. Images with different focus far and close.

3. Light field imaging

Light field system inherently measure projections of the four dimensional (4d) light field scalar function onto a two dimensional sensor and therefore, suffer from a spatial vs. Angular resolution trade ff. programmable light field imagers, proposed recently, overcome this spatio angular resolution trade off and allow high resolution capture of the (4d) light field function with multiple measurements at the cost of a longer exposure time.

The plenoptic function, as defined by adelson and berger, allows the reconstruction of every possible view, at every moment, from every position, at every wavelength, within the bounds of the space-time-wavelength region under consideration [ADE91].

$$L = L(x, y, \lambda, t, u, v, w) \tag{1.2}$$

In the above equation, l is the light field, (u, v, w) represents location of the camera aperture in every possible location, (x, y) represents every location on the image sensor, λ represents every wavelengths and t represents every time. When considering only the still imagery, plenoptic function can be defined as a five dimensional (5d) function.

$$L = L(x, y, u, v, w) \tag{1.3}$$

Moon and spencer, and Levoy and Hanrahan[LEV96]found redundancy in the above equation when the (u, v) plane has a fixed location and reduced the 5D plenoptic function to a 4 dimensional (4D) function.

$$L = L(x, y, u, v) \tag{1.4}$$

The 4D plenoptic function in Equation (1.4) is referred to with different names: photic field, 4Dlumigraph [GOR96]or 4D light field[LEV96].From here on, the use of the plenoptic function refers to the 4D plenoptic function given in equation (1.4).

4. Plenoptic camera

Images or light field can be captured in a variety of ways [BOL14]:

- ❖ By mounting a camera on a gantry and taking images from different positions,
- ❖ By using an array of cameras, or
- ❖ By mounting a microlens array between the lens and the image sensor of the camera.

In this work, plenoptic cameras with microlens arrays mounted between the lens and the image sensor are used. A schematic of the plenoptic camera used in this research is shown in Figure (1.3).

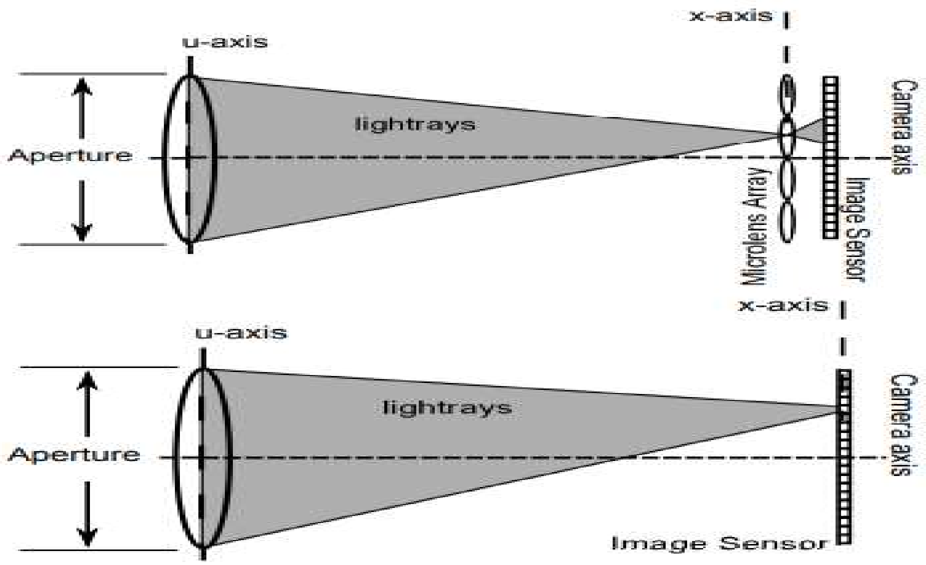


Figure1.3. Images showing the schematic of (a) plenoptic and (b) conventional cameras.

The difference in the construction of the plenoptic camera from a conventional camera causes the difference in the way an image is formed on their image sensors.

In a conventional camera, shown in figure (1.3), light rays passing through the lens aperture converge on a pixel and the pixel value is a function of the weighted sum of light converging from the aperture. The weight here is a function of the angle between the incident ray and the normal of the image sensor plane. In integrating the light coming from the aperture, all the directional information of the light rays is lost. In plenoptic cameras, the use of the microlens array divides the image sensor into sub regions of equal size.

The light coming from the lens aperture converges on the microlens array as opposed to the conventional camera where the light converges on the image sensor. The image formed by each microlens is essentially an image of the main lens aperture such that each pixel in the sub region corresponds to a different point on the aperture, thus preserving the angular information of the light field. In figure (1.4), the idea of storing the information in a pixel from a particular region of the lens in the plenoptic cameras is seen.

Extending the idea to demonstrate the effect and the difference in the image formation between a conventional and the plenoptic cameras, schematic images showing image formation when a light source is:

- ❖ At the focal plane.
- ❖ Located closer to the camera than the focal plane.
- ❖ Located farther away from the camera than the focal plane.

With the conventional camera, when the light source is at the focal plane of the camera setup, a sharp image is formed on the image sensor. Moving the light source closer or farther from the camera blurs the captured image. The two blurred images of the light show no difference when compared to each other. When imaging the same light source with the plenoptic camera, the captured images in all three instances difference from one another. When the light source is placed at the focal plane, the light source's image fills up a microlens on the microlens array and the pixels right behind it on the image sensor.

As the source is moved away from the focal plane, either closer to or away from the camera, the image of the light source on the microlens array blurs and spreads to more than one microlens. As the incident angle of the light on the microlens array varies depending on whether it is moved closer to or farther away from the camera, only certain pixels under the microlens get illuminated as a function of the incident angle. Illumination of the pixels behind the microlens as a function of the incident angle varies the image formed on the image sensor. Thus the information captured in plenoptic cameras carry both spatial and angular information of the captured light.

5. Light field parameterization

Some alternative parameterizations of the 4d light field, which represents the flow of light through an empty region of three dimensional space. Left points on a plane or curved surface and directions leaving each point center: pairs of points on the surface of a sphere right: pairs of points on two planes in general (meaning any position).

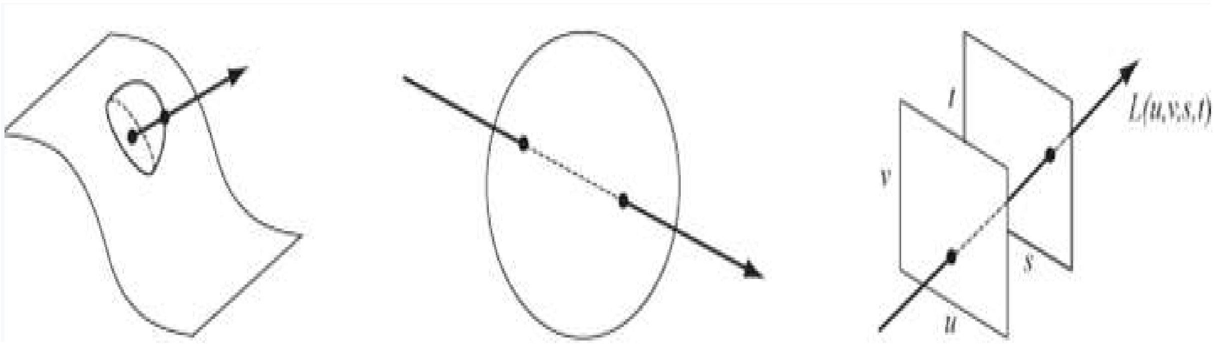


Figure1.4.light field parameterization.

5.1 Tow planes parameterization

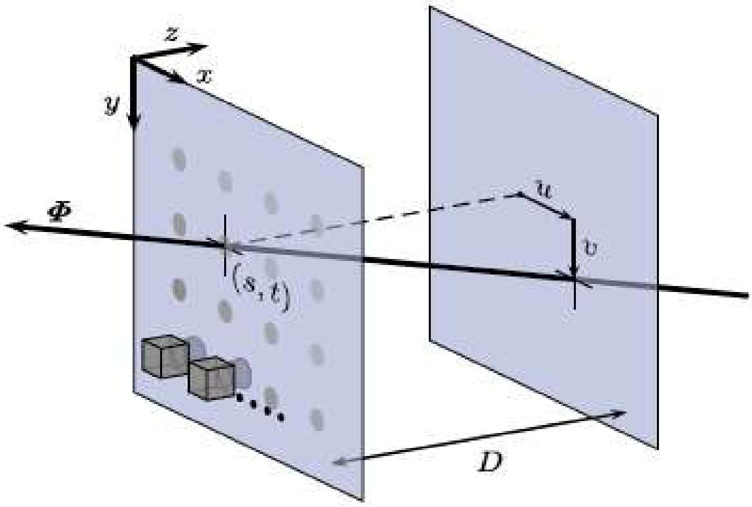


Figure 1.5. Two-plane parameterizations of light rays – shown are the relative two-plane parameterization.

The point of intersection of a ray with two parallel planes completely describes its position and orientation in space. By convention, the s, t plane is closer to the camera, and the, v plane is closer to the scene.

This realization allows light rays to be parameterized in terms of four dimensions instead of five. A common way of doing so measures each ray's points of intersection with two parallel reference planes, requiring only four numbers to describe the ray, two for each of position and direction.

That the light rays not impact a surface or pass through attenuating media are not as restrictive as they may first appear. This means, for example, that a light field model of a closed door will allow rendering of novel views in front, throughout this work we employ light field parameterizations in which light rays are described by their points of intersection with two parallel planes: an s, t plane, by convention closest to the camera, and a u, v plane at distance d, by convention closer to the scene.

The continuous-domain light field signal L(s, t, u,v) describes all light rays passing through. The spherical parameterization more easily suggests the fact that the light impinges on a given point in space from all directions and that no direction has special status. However, the Cartesian parameterization.

6. Light field acquisition

This section focuses on existing devices or methods for light field acquisition. A conventional camera captures a 2D projection of a light field on its sensor plane by integrating the light rays that hit each pixel from every direction. In contrast, the devices or methods for light field acquisition measure the distribution of light rays in a directionally resolved manner, avoiding angular integration.

However, sensors can only measure the information from two dimensions (usually two spatial dimensions) of a scene at a single moment. To acquire a 4D light field, we need to capture multiple samples along the angular dimensions. Existing light field acquisition approaches can be divided into three fundamental categories multi-sensor capture, time-sequential capture and multiplexed imaging. In the following, we will provide a detailed review of the existing light field acquisition approaches according to those three categories [LEV 06] [IHR 16] [WET11] [IHR10].

6.1. Multi-sensor capture

The multi-sensor capture approach requires an array of image sensors distributed on a planar or spherical surface to simultaneously capture light field samples from different viewpoints [LEV 96].The spatial dimensions (u and v) of the light field are determined by the sensors, while the angular dimensions (s and t) are determined by the number of cameras and their distribution.

Therefore, the 4D light field is recorded by the combination of the captured images. In 2002, Yang et al.[YAN02] Described a design using an array of 88 video cameras for dynamic light field capture.

To overcome data bandwidth problems, they employed distributed rendering algorithm. In the same year, wilburnetal [WIL01].Applied 6 CMOS image sensors to record synchronized video dataset.

- ❖ A camera array system composed of 8×12 video cameras developed by Wilburn et al.
- ❖ The Picam developed by Venkataraman and always an ultra-thin monolithic 4×4 camera array that could be integrated into a cell phone.
- ❖ Lin et alused a 5×5 camera array system to capture the light field of microscopic objects

Figure1.6. Devices for multi-sensor capture.

6.2 .Times-sequential captures

In contrast with the multi-sensor approach, a time-sequential capture approach uses a single image sensor to capture multiple samples of the light field through multiple exposures. The typical approach uses a sensor mounted on a mechanical gantry to measure the light field at different positions [LEV 96].

Two gantry systems were presented by the Computer Graphics Laboratory at Stanford University [LIP08] one was a computer-controlled gantry with four degrees of freedom, translation in X and Y, nod and shake; another was a Lego Mind storms gantry in which the motors have rotator controllers enabling the camera to move along accurate and well-defined paths. Unger et al. [UNG03]used a single camera mounted on a motorized linear stage with two degrees of freedom, translation in X and Y, to capture light fields.

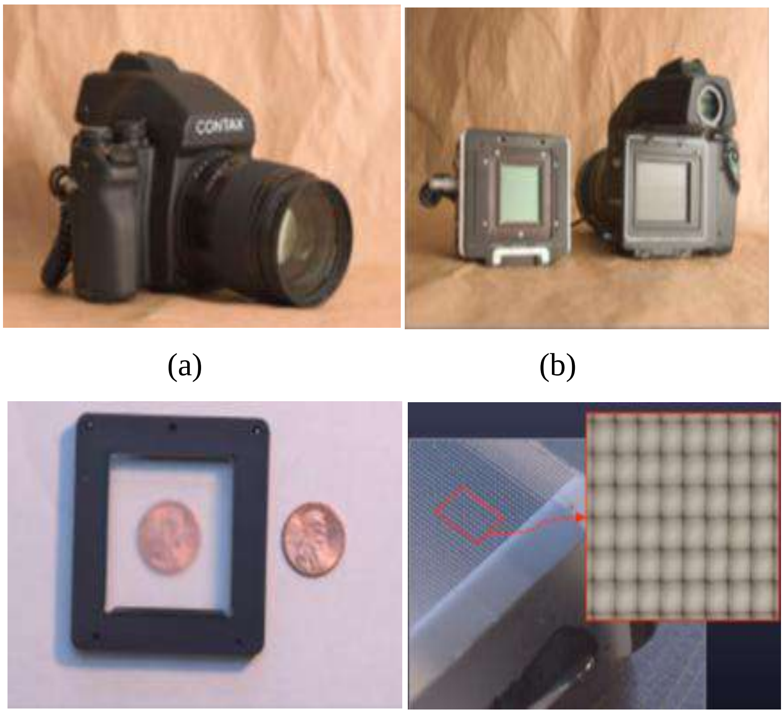
Figure1.7.”4 degree- of- freedom gantry”Image Levoy et al.

6.3. Multiplexed imaging

The last approach aims to encode the 4D light field into a 2D sensor plane, by multiplexing the angular domain into the spatial (or frequency) domain. It allows dynamic light field capture with a single image sensor, but imposes a tradeoff between the spatial and angular resolutions (one can obtain densely sampled images in the spatial domain with sparse samples in the angular domain, and vice versa).Multiplexed imaging can be further divided into spatial multiplexing and frequency multiplexing.

6.3.1. Spatial multiplexing

In spatial multiplexing, an interlaced array of elemental images representing samples from different 2D slices of the light field is captured by the sensor. Most spatial multiplexing approaches are implemented using an array of microlenses or a lens let array mounted on the image sensor. Interestingly, this is one of the first approaches for light field imaging: in 1908, Lippmann used this approach in the development of his “integral photography” [LIP08].



(a) (b)

(c) (d)

Figure 1.8. Prototype camera , (a) Medium format digital camera, (b) 16 megapixel sensor, (c) (d) microlens array

6.3.2 .Frequency multiplexing

Unlike the spatial multiplexing approach, which interlaces the 2D light field slices on the sensor plane, the frequency multiplexing approach encodes different 2D slices of the light field into different frequency bands. Typically, frequency multiplexing approaches use a modulation mask to achieve a certain property in the Fourier domain [IHR10] .Veeraraghavan et al. [AGR10] described a theoretical framework for 4D light field acquisition using an attenuation mask in the optical path of a conventional image sensor; they termed the approach "Dappled Photography". Rather than blending the light rays, the patterned mask attenuated and encoded them on the image sensor. The Fourier transformed image is rearranged into 4D planes and then an inverse Fourier transform is applied to restore the light field.

Figure1.9. Prototype camera design Encoded blur camera holds a coarse broadband mask (shown in right) in the lens aperture.

7. Imaging model

The light field camera is approximately equivalent to a camera with two types of optical elements: a main lens and a microlens array. As in [LUM08], we consider the imaging system under a general configuration of these optical elements; however, unlike in any previous work, we determine the image formation model of the camera so that it can be used for.

We summarize all the symbols and their meaning in Table 1.1 As shown in Figure (1.11), the image of a 3D point in space.

$$p = [xyz]^{T} \in \mathbb{R}^3 \tag{1.5}$$

Results in a collection of blur discs whose shape depends on three factors: the blur introduced by the main lens, the masking due to the micro lens array, and the blur introduced by the micro lenses. We will see all of these effects in the following analysis and, in particular, in section by applying the thin lens law I to the main lens, we can find that the point p is brought into focus inside the camera at the position [LUM08].

$$p' = [x'y'z']^{T} \in \mathbb{R}^3 \tag{1.6}$$

The plenoptic camera comprises a main photographic lens, a microlens array, and a digital photosensor. The microlenses are drawn the image artificially large to make it possible to see them and the overall camera at the same scale. In reality they are microscopic compared to the main lens, plane is the imaging plane, and the size of the individual microlenses sets the spatial sampling resolution.

Table 1.1: Light field camera symbols and their description.

	Camera parameters
D	Main Lens Diameter
d	Microlens Diameter
F	Main lens focal length
f	Microlens focal length
c	Microlens center in 3D
v'	Microlens to CCD sensor distance
u	Size of CCD sensor element
	Scene parameters
P	3D point in space
p'	Focused image of p inside the camera
P''	Projection of p onto the CCD sensor
	Point spread function parameters
P^B	Main lens blur center in 3d space
B	Main lens blur radius
b	Microlens blur radius

All the light that passes through a pixel passes through its parent microlens and through its conjugate square (sub aperture) on the main lens. Bottom: all rays passing through the sub-aperture are focused through corresponding pixels under different microlenses. These pixels form the photograph seen through this sub.

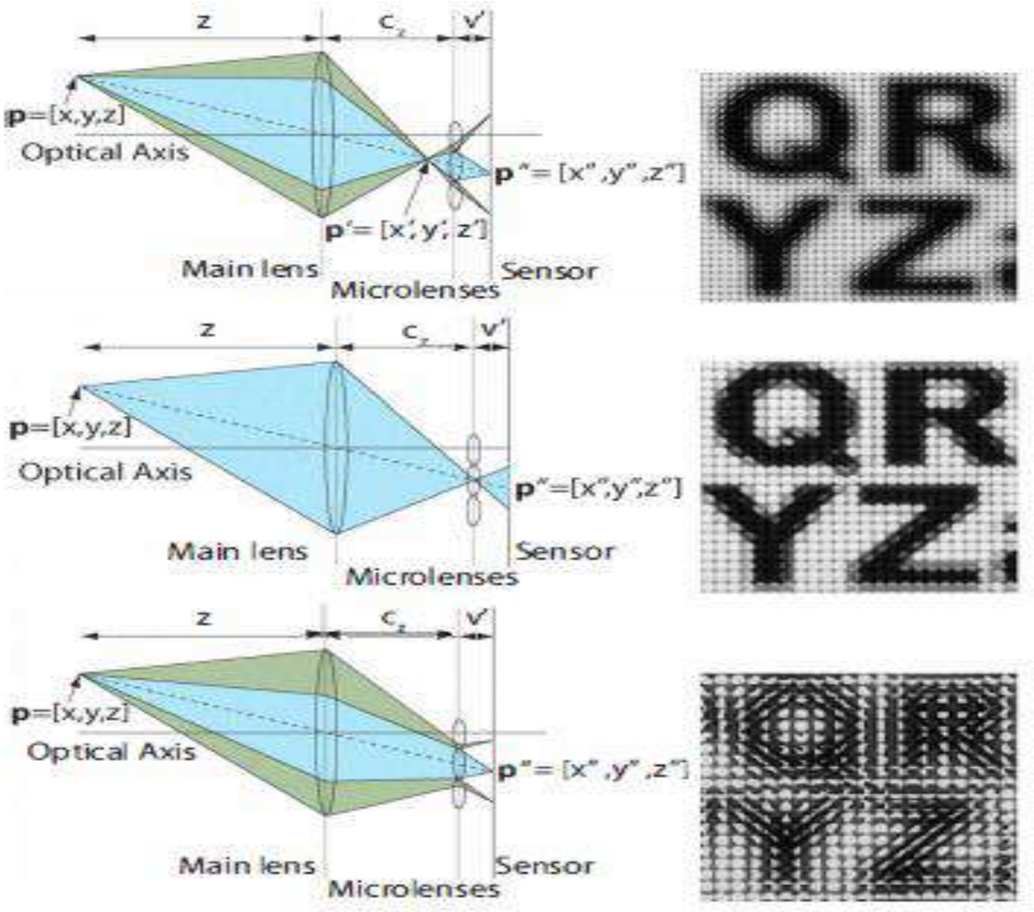


Figure 1.10: Schematic of a 2D section of a light field camera.

8. Applications

Given the wide variety of light field capture devices, a similarly diverse set of applications is enabled by such high-dimensional representations of light transport. Light fields have proven useful for large variety of applications in computer graphics, digital photography, and 3dreconstruction. Specially, these include image-based rendering. [GOR96]

Features that light field cameras offer are: post-capture refocus, change of view point, three-dimensional (3d) data extraction, change of focal length, focusing through occludes, increasing visibility in bad weather conditions, and improving the robustness of robot navigation, to name just a few.

9. Conclusion

This chapter describes first the conventional imaging paradigm and presents its limits. Here we focus on the problem of defocus. Then, we present the new technology of light field imaging of which enables capturing full light information in a scene. We explained the concept of the plenoptic function and the parameterisation of the light field. Light field acquisition can be done by many techniques; we presented here of them: the multi sensor capture, time sequential, capture multiplexed, imaging spatial multiplexing, and frequency multiplexing lf. So we can say that in a light field imaging, photos are generated after the acquisition of the light field, then the user can modify (change focus, make a rotation...) photos before displaying. In the next chapter, we present in details how one can change the focus in the light field imaging.

Chapter 2

Images from the light field and digital refocusing

1. Introduction

Recent developments in computational photography enabled variation of the optical focus of a plenoptic camera after image exposure also known as refocusing and now it is possible to fix the photo even after it is captured, In this Chapter, we will present the image formation next we will see Plenoptic Camera Records the Light Field and the Imaging Equations and also Digital Refocusing with some examples, finally the Image Synthesis Algorithms are presented.

2. Image formation

In a conventional camera, a photograph forms on a piece of photosensitive material placed inside the camera at the imaging plane. Where photons cause the development of silver crystals, or a CCD or CMOS photo sensor in a digital camera, where photons generate free electrons that accumulate in each sensor pixel. Each position on the photosensitive imaging plane sums all the rays of light that terminate there.

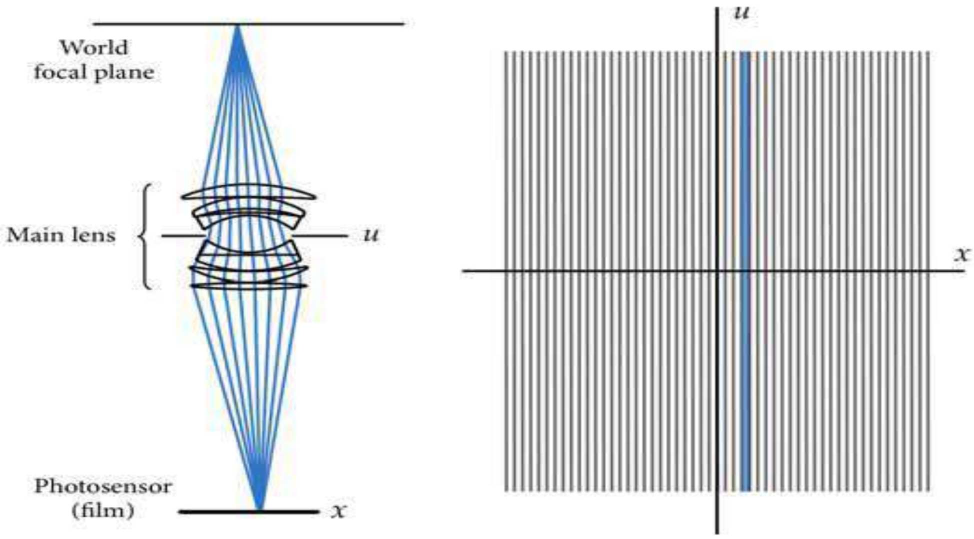
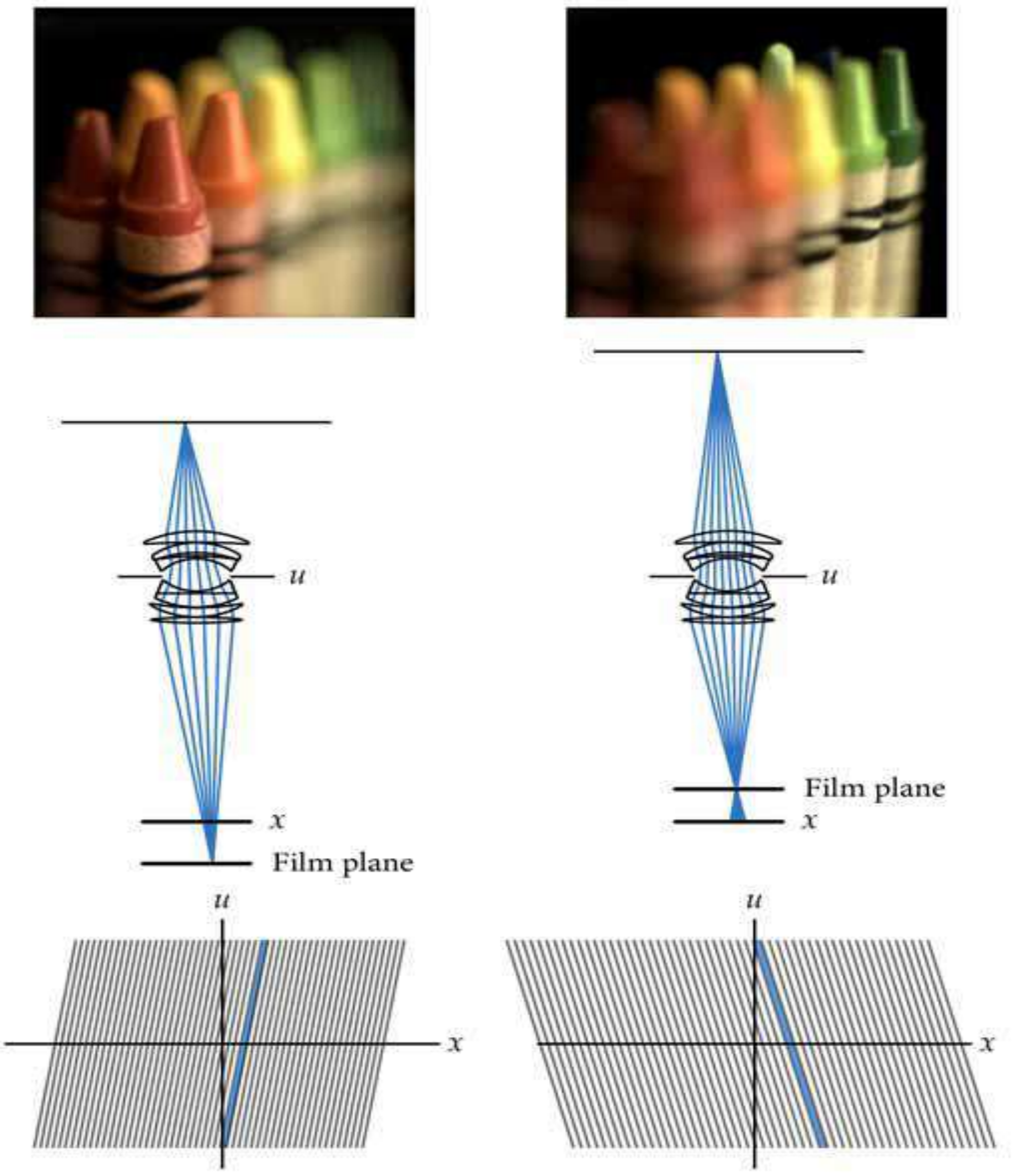


Figure 2.1. Draws in blue the cone of rays contributing to one photograph pixel value.

This cone corresponds (in 2d) to the blue vertical strip on the ray-space diagram because the rays in the cone share the same x film intercept, but vary over all u positions on the lens. Of course different pixels in the photograph have different x intercepts, so they correspond to different vertical lines on the ray-space. In fact, the ray-space drawn in figure (2.1) is overlaid with vertical strips, where each strip is the set of rays summed by a different photograph pixel. This drawing shows that the formation of a full photograph corresponds on the ray-space diagram to a vertical projection of the light field values. The projection preserves the spatial x location of the rays, but destroys the directional information.



(A) (B)

Figure 2.2. Illustrates how the projection of the ray space changes as the camera is focused at different depths.

Figure (2.2) illustrates how the projection of the ray space changes as the camera is focused at different depths. In these diagrams, the x plane is held fixed at a canonical depth while the film plane of the camera moves. Changing the separation between the film and the lens is how we focus at different depths in a conventional camera.

For example in Figure (2.2) A, the film plane is moved further from the lens, and the world focal plane moves closer to the camera. The cone of blue rays corresponds to the blue strip with positive slope on the ray-diagram.

As the intercept of a ray moves linearly across the u lens plane, the resulting intercept moves linearly across the x film plane. If the convergence point is further from the lens than the x plane, then the movement across the u plane is in the same direction as the movement across the x plane Figure (2.2) A. These two directions are opposed if the convergence point is in front of the x plane Figure (2.2) B. These figures make it visually clear that the relative rates of the movements, hence slopes on the ray-space diagram, depend on the separation between the convergence point and the x plane.

3. Plenoptic camera records the fight field

A grid of boxes lies over the ray-space diagram in Figure (2.3) this grid depicts the sampling of the light field recorded by the photosensor pixels, where each box represents the bundle of rays contributing to one pixel on the photosensor. To compute the sampling grid, rays were traced from the boundary of each photosensor pixel out into the world through its parent microlens array and through the glass elements of the main lens.

The intercept of the ray with the microlens plane and the lens plane determined its (x, u) position on the ray-space diagram. As an aside, one may notice that the lines in the ray-space diagram are not perfectly straight, but slightly curved. The vertical curvature is due to aberrations in the optical design of the lens.

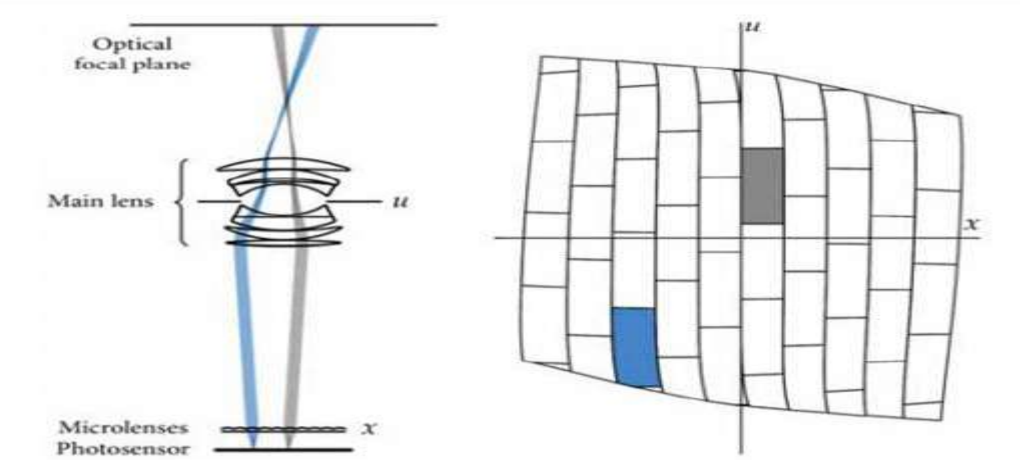


Figure 2.3. Sampling of a photograph's light field provided by a plenoptic camera

Each sample box on the ray-space corresponds to a narrow bundle of rays inside the camera. For example, figure (2.3) depicts two colored sample boxes, with corresponding ray-bundles on the ray diagram. A column of ray-space boxes corresponds to the set of all rays striking a microlens, which are optically sorted by direction onto the pixels underneath the microlens, as shown by the gray boxes and rays in figure (2.3) if we summed the gray photosensor pixel samples on figure (2.3), we would compute the value of an output pixel the size of the microlens, in a photograph that were focused on the optical focal plane. These examples highlight how different the plenoptic_ camera sampling grid is compared to that for a conventional camera, in the conventional camera, all the grid cells extend from the top to the bottom of the ray space, and their corresponding set of rays in the camera is a cone that subtends the entire aperture of the lens.

In the conventional camera the width of a grid column is the width of a photosensor pixel. In the plenoptic camera, on the other hand, the grid cells are shorter and wider. The column width is the width of a microlens, and the column is vertically divided into the number of pixels across the width of the microlens. In other words, the plenoptic camera sampling grid provides more specificity in the u directional axis but less specificity in the x spatial axis, assuming aconstant number of photosensor pixels.

This is the fundamental trade-off taken by the light field approach to imaging. For a fixed sensor resolution, collecting directional resolution results in lower resolution final images, with essentially as many pixels as microlenses. On the other hand, using a higher-resolution sensor allows us to add directional resolution by collecting more data, without necessarily sacrificing final image resolution.

4. Imaging equations

This mathematical relationship is a natural basis for computing photographs focused at different depths from the light fields recorded by the camera

$$E_F(x,y) = \frac{1}{F^2} \iint L_F(x,y,u,v) \cos^4\theta \text{ du dv} \tag{2.1}$$

Where F is the separation between the lens and the film, EF(x, y) is the irradiance on the film at position (x, y), LF is the light field parameterized by the planes at separation F, and θ is the angle between ray (x, y, u, v) and the film plane normal. For simplicity, Equation (2.1) also assumes that the u v and x y planes are infinite in extent, and that L is simply zero beyond the physical bounds of the lens and sensor. To further simplify the equations in the derivations throughout the thesis, let us also absorb the $cos^4\theta$ into the definition of the light field itself, by re-defining L(x, y, u, v) **:** = L(x, y, u, v) $cos^4\theta$.

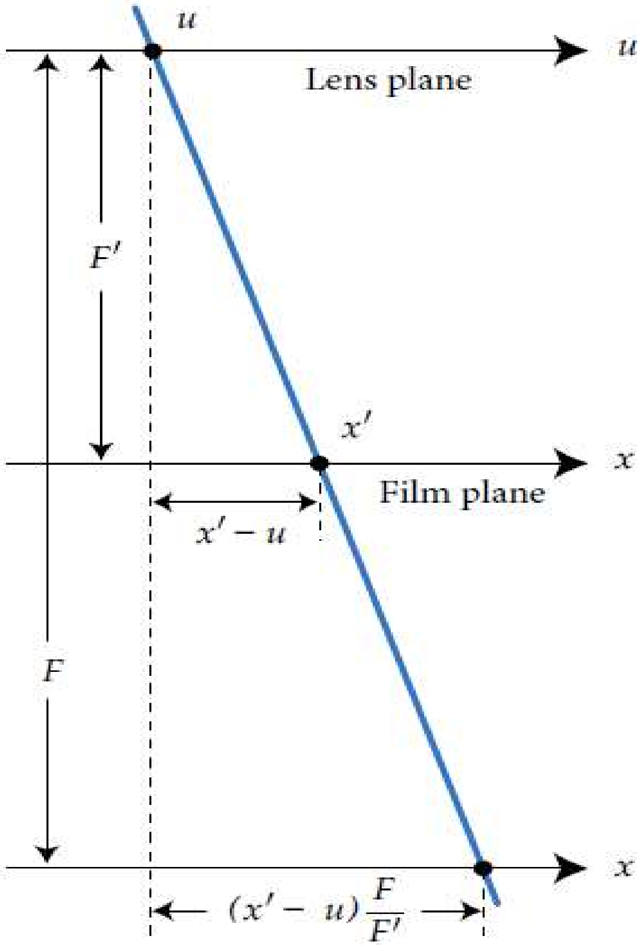


Figure 2.4. Transforming ray-space coordinates.

The photographs focused at depths other than the x parameterization plane. As shown in Figure (2.4), focusing at different depths corresponds to changing the separation between the lens and the film plane, resulting in a shearing of the trajectory of the integration lines on the ray-space. If we consider the photograph focused at a new film depth of F’, then deriving its imaging equation is a matter of expressing $L_{F\prime}$(x’, u) in terms of L_F(x, u) and then applying Equation (2.1).

The diagram above is a geometric construction that illustrates how a ray parameterized by the x’ and u planes for $L_{F\prime}$may be re-parameterized by its intersection with planes x and u for L_F. By similar triangles, the illustrated ray that intersects the lens at u and the film plane at x’, also intersect, the x plane at u + (x’− u) F/F’. Although the diagram only shows the 2d case involving x and u, the y and v dimensions share an identical relationship. As a result, if we define α = F’/F as the relative depth of the film plane [HAN06].

$$L_{F'}(x',y',u,v) = L_F(u + \frac{x'-u}{\alpha}, v + \frac{y'-v}{\alpha}, u, v)$$

$$= L_F(u\left(1-\frac{1}{\alpha}\right) + \frac{x\prime}{\alpha\prime}, v\left(1-\frac{1}{\alpha}\right) + \left(\frac{y'}{\alpha}, u, v\right). \quad (2.2)$$

This equation formalizes the 4d shear of the canonical light field that results from focusing at different depths. Combining equations (2.1) and (2.2) leads to the final equation for the pixel value (x, y) in the photograph focused on a piece of film at depth F= α **.** F from the lens plane.

For example, figure (2.5) a, illustrates in blue all the rays that contribute to one pixel in a photograph refocused on the indicated virtual focal plane., this blue cone corresponds to a slanted blue line on the ray-space, as shown in figure (2.5). b1, to synthesize the value of the pixel, we would estimate the integral of light rays on this slanted line. The second concept is that the radiance along rays in the world can be found by ray- tracing.

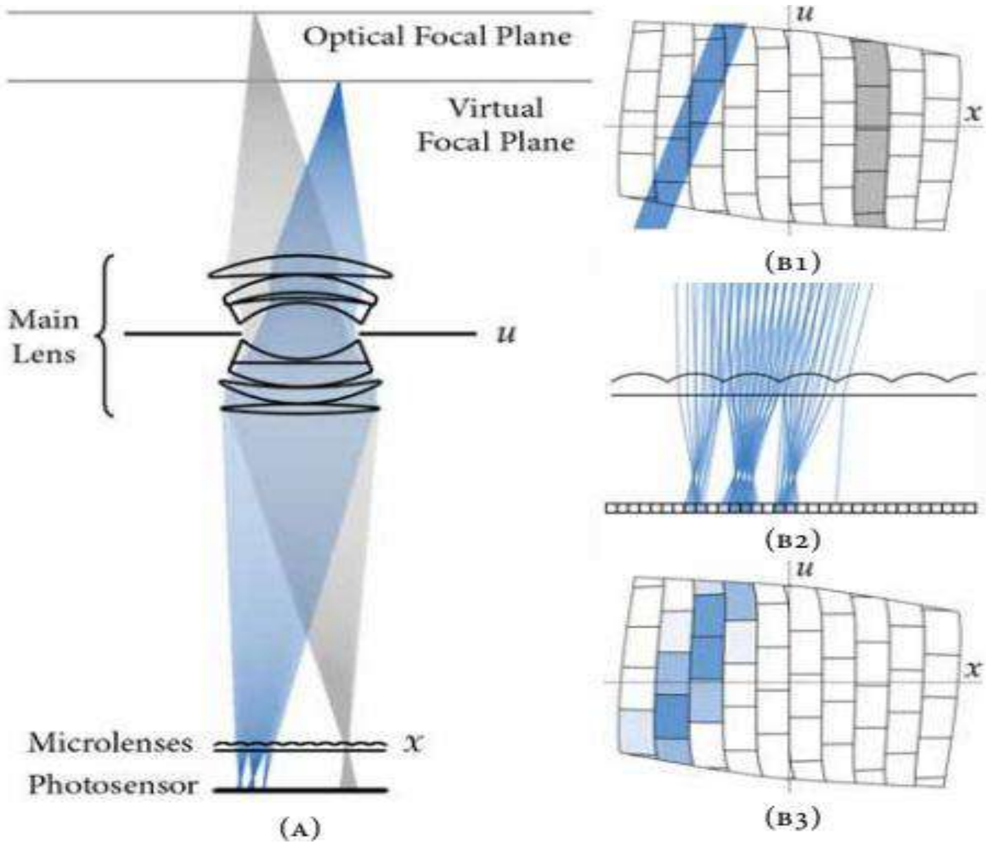


Figure2.5. Overview of processing the recorded light field

Specifically, to find the radiance along the rays in the slanted blue strip, we would geometrically trace the ideal rays from the world through the main lens optics, through the microlens array, down to the photosensor surface. This process is illustrated macroscopic-cally by the blue rays in Figure (2.5) a. Figure (2.5) b2 illustrates a close-up of the rays tracing through the microlenses down to the photosensor surface.

Each of the shaded sensor pixels corresponds to a shaded box on the ray-space of figure (2.5) b3 weighting and summing these boxes estimates the ideal blue strip on figure (2.5) b1 the number of rays striking each photosensor pixel in b2 determines its weight in the integral estimate. This process can be thought of as rasterizing the strip onto the ray-space grid and summing the rastered pixels. There are more efficient ways to estimate the integral for the relatively simple case of refocusing, and optimizations are presented in the next chapter.

However, the ray-tracing concept is very general and subsumes any camera configuration. For example, it handles the case of the generalized light field camera model where the micro lenses defocus [HAN06].

5. Digital refocusing

People who see these images for the first time are often surprised at the high fidelity of such digital refocusing from light fields. Images a1-a5look like the images that I saw in the view finder of the camera as I turned the focus ring to focus on the girl in the white cap. The underlying reason for this fidelity is that digital refocusing is based on a physical simulation of the way photographs from inside a real camera.

(A1) (A2) (A3)

(A) (b)

Figure 2.6. Examples of refocusing (A1-A4) and extended depth of field (b)

Figure (2.6) b is a different kind of computed photograph, illustrating digitally extended depth of field. In this image, every person in the scene is in focus at the same time. This is the image that a conventional camera would have produced if we had reduced the size of the lens aperture in the classical photographic method of optically extending the depth of field. Image b was computed by combining the sharpest portions of images a1-a5 [Agarwala et al. 2004], and can be thought of as refocusing each pixel at the depth of the closest object in that direction [HAN06].

6. Image synthesis algorithms

The ideal set of rays contributing to a pixel in digital refocusing is the set of rays that converge on that pixel in a virtual conventional camera focused at the desired depth, which formally specifies in its integral the set of light rays for the pixel at position (x, y):

$$E(\alpha.F)(x',y') = 1/\alpha^2F^2 \iint Lf(u(1-\tfrac{1}{\alpha}) + \tfrac{x'}{\alpha}, v\left(1-\tfrac{1}{\alpha}\right) + \tfrac{y'}{\alpha}, u, v)\, dudv. \quad (2.3)$$

Recall that in this equation LF is the light field parameterized by an x y plane at a depth of F from the u v lens plane, α is the depth of the virtual film plane relative to F, and E (α F) is the photograph formed on virtual film at a depth of (α F) One way to evaluate this integral is to apply numerical quadrature techniques, such as sampling the integrand for different values of u and v and summing them. The idea is to trace the ray $(u\left(1-\frac{1}{\alpha}\right)+\frac{x'}{\alpha}, v\left(1-\frac{1}{\alpha}\right)+\frac{y'}{\alpha}, u\,, v)$,through the microlens array and down to the photosensor. The intersection point is where the ray deposited its energy in the camera during the exposure, and the value of LF is estimated from the photosensor values near this point.

However, a more efficient method is suggested by the linearity of the integral with respect to the underlying light field. Reveals the important observation that refocusing is conceptually a summation of dilated and shifted versions of the sub-aperture images over the entire u v aperture. This point is made clearer by explicitly defining the sub-aperture image at lens position (u, v) in the light field LF. Let us represent this sub-aperture image by the 2d function $lf^{(u,v)}$, such that the pixel at position (x, y) in the sub-aperture image is given by$lf^{(u,v)}(x,y)$.

With this notation, we can re-write Equation (2.3) :

$$E_{(\alpha.F)}(x',y') = \frac{1}{\alpha^2F^2}\iint LF^{\ (u,v)}\left(u\left(1-\frac{1}{\alpha}\right)+\frac{x\prime}{\alpha}, v\left(1-\frac{1}{\alpha}\right)+\frac{y\prime}{\alpha}\right) dudv \quad (2.4)$$

where $lf^{(u,v)}(u\left(1-\frac{1}{\alpha}\right)+\frac{x}{\alpha}, v\left(1-\frac{1}{\alpha}\right)+\frac{y}{\alpha}$

Is simply the sub-aperture image L (u,v) F , dilated by a factor of α and shifted by a factor of $(u\left(1-\frac{1}{\alpha}\right), v\left(1-\frac{1}{\alpha}\right))$. In other words, digital refocusing can be implemented by shifting and adding the sub-aperture images of the light field. This technique has been applied in related work on synthetic aperture imaging using light fields acquired with an array of cameras.

7. Conclusion

In this chapter we presented the image formation and the digital refocusing plenoptic camera records the light field and the imaging equations also we saw examples of refocusing and extended depth of field and finally we take a overview to the image synthesis algorithms.

Chapter 3

Face detection and image fusion

1. Introduction

In video surveillance, face detection is essential. Several techniques have been proposed for face detection. In the first part of this chapter we give an overview of the proposed methods. We focus on the method of fusion .One of the problems of face detection is in the case where the person appears in a camera out of focus plan. This makes the face (region of interest) blurry. To remedy this problem, an acquisition of several images with different focus is necessary. The fusion of its images gives a clearer and more practical image in the detection. Thus, the second part of this chapter will present the fusion of images. We start by giving the merger levels then we perform some fusion techniques, those we will use later. We also show the evaluation parameters of the marger. The detection of faces that are out of focus requires the acquisition of multiple images.

2. Face Detection

The definition of face detection refers to a subset of computer technology that is able to identify people’s faces within digital images. Face detection applications employ algorithms focused on detecting human faces within larger images that might contain landscapes, objects and other parts of humans. Facial recognition is a category of biometric software that maps an individual's facial features mathematically and stores the data as a face print. One of the most important techniques that enable human-computer interaction is face detection. In general facial analysis algorithms such as face recognition alignment, facial expression tracking/recognition, age estimation, and head pose tracking and many more, face detection is the first and key step as shown in.

Figure 3.1. Automatic face verification system

3. Some methods of face detection

Face detection is to locate faces in a image. It can be seen as pre-processing of face recognition/verification/identification. this is some methods of face detection.

3.1. Knowledge-based methods

Face detection technology might begin by searching for human eyes. It might do this by testing valley regions in the gray-level image. It might then use a genetic algorithm to detect facial regions including eyebrows, the mouth, nose, nostrils and the iris. The algorithm would first identify possible facial regions and then apply additional testing in order to validate.

A first way to detect faces on image is using a set of simple rules that describe with accuracy the human-face proportions from facial images. The rules are provided by a human expert, for example, the center of a facial image has uniform intensity values as well as a considerable intensity difference with the borders and also some morphological considerations like a human-face has always two eyes, a noise and a mouth.

3.2. Feature invariant methods

The main objective of this method is to find a set of facial structure features invariant to pose and lighting variation it detect the most important candidate region using a neural network trained to recognize human skin color, finally a Gaussian model then is applied to chose the regions that correspond to a face. While feature invariant methods show an invariance to pose, methods based on skin color are affected by changes in illumination color and intensity, image noise and occlusions. Methods based in facial edges can be affected by shadows, giving some non desired facial edges.

3.3. Template matching methods

In Template Matching methods a standard face-template is correlated with candidate regions, and correlation scores are computed individually for each facial component (nose, mouth, eyes etc). A candidate region is validated as a face by aggregating their scores. In spite of the simplicity of this type of method for detection, a good definition of template remains challenging in the case of multi-view face detection because a template is necessary for each face pose.

3.4. Appearance-based methods

As was observed in the preceding sections, face detection methods, the definition of a face model is a challenging task that sometimes requires the support of a human expert to define a set of rules or a template that correctly models human facial variations. In this scope, appearance-based methods try to capture the most representative variations from a set of example images. To capture such variations in an automatic way without human-intervention, researchers in this field have used machine learning and statistical analysis methods such as Principal Component Analysis, Neural Networks.

4. Eigen faces algorithm

The facial recognition method Eigen faces employs the technique of principal component analysis, which marks a notable difference with the more traditional methods, called local methods, which are based on the peculiarities of the analyzed face, and whose defects lie in its lack. Accuracy, as well as its sensitivity to information that is irrelevant. The method we used is termed global, since the whole face is then analyzed. Our recognition technique will therefore use the principal component analysis method (also called PCA). In a simple way, we aim to reduce the size of the space in which we.

We will work, and we can then simplify the data at our disposal and their interpretation. The ACAP algorithm, PCA English (Principal Component Analysis) was developed by M.A. Turk and A.P. Pent land at MIT Media Lab in 1991 [TUR91].It is also known as Eigen faces because it uses eigenvectors and eigen values. This algorithm relies on well-known statistical properties and uses linear algebra. It is relatively quick to implement but remains sensitive to problems of illumination [TUR91], pose and facial expression.

The main idea is to express the M learning images according to a base of particular orthogonal vectors, containing independent information from one vector to another. These new data are therefore expressed in a more appropriate way for facial recognition. We want to extract the characteristic information of a face image, to encode it as efficiently as possible in order to compare it to a database of similarly encoded models. In mathematical terms, this amounts to finding the eigenvectors of the covariance matrix formed by the different images of our learning base [TOL06]. We will seek to find clean faces; first, we must take a number M of learning faces. Each of these images, which are in practice $N \times N$matrices, are then transformed into a single column vector of length$N2$. Initial $N \times N$matrix:

$$\begin{pmatrix} \alpha 11 & \alpha 12 & \alpha 1i & \alpha 1n \\ \alpha 21 & \alpha 22 & \alpha 2i & \alpha 2n \\ \alpha n1 & \alpha n2 & \alpha ni & \alpha nn \end{pmatrix}$$

Transformed into:

$$\begin{pmatrix} \alpha 11 \\ \vdots \\ \alpha n1 \\ \vdots \\ \alpha 1n \\ \vdots \\ \alpha nn \end{pmatrix}$$

We must then determine the average face, deduced from the M faces of learning.

$$\Psi = 1/\mathrm{M} \sum_{i=1}^{M} Ti \tag{3.1}$$

Figure 3.2. Middle face.

This average face will be used in the analysis of images, we subtract effect and this average face faces of learning, which leaves us the information specific to this face, we then recover in only lesi information that is specific to this learning face.

$$\varphi i = \Gamma\mathrm{i} - \Psi \tag{3.2}$$

Where φi represents the I eme face to which we subtracted the average face. Now we have to calculate the covariance matrix D. It corresponds to:

$$D = QQ^T \tag{3.3}$$

With $$\mathrm{Q} = [\Phi 1, \Phi 2, \cdots, \Phi \mathrm{M}] \tag{3.4}$$

We should calculate the eigenvectors di of the matrix D. But this represents for us

N eigenvectors of dimension $N2$ each. It is now that we will reduce the information by limiting the components we will work with, in accordance with the principle of principal component analysis. We will therefore consider the matrix $E = Q^T Q$, we will find eigenvectors ei. This matrix is of size $M \times M$ which will therefore simplify things since we will have M eigenvectors of size M each. The transition from the matrix D to the matrix E is not trivial, we use the fact that the eigenvectors of these two matrices are linked rather closely. In fact, we have as a relation.

$$Ee_i = Q^T Qei = \lambda_i e_i \tag{3.5}$$

With λi the eigen value associated with the Eigen vector ei. By multiplying this equation by the matrix$Q$, it comes.

$$QEe_i = QQ^T Qe_i \tag{3.6}$$

We then see the matrix D.

$$QEei = DQei = \lambda i Qei \tag{3.7}$$

We thus deduce that with eieigenvector of the matrix E associated with the eigen value λi, we therefore have Qei is an eigenvector of the matrix D associated with the same eigen value λi. So, we have di e own vector of D, with

$$di = Qei \tag{3.8}$$

These are the proper values that they associate with each other, so that we can remove the eigenvectors according to their ability to characterize the variations between the images. When they are visualized (these vectors are at the origin of the N × N squares), the eigen faces are what might be called ghostly images. But keep in mind that these are the eigenvectors of the covariance matrix of face learning images.

Figure 3.3. Some clean faces.

The M eigenvectors we then obtained will allow us to better approximate the faces of learning by using the eigen faces of greater importance. The advantage of reducing the number of clean faces is on the one hand require less memory space, but also reduce calculations, their execution time; however, we are undoubtedly losing information and therefore less accurate information, but the results will not really change, given that we are giving ourselves only one identification mission. We do not seek to reconstruct the face of the subject from our own faces, but only to recognize it. Among the M eigenvectors found, we will only keep a number L, which will be the most significant. We will now find the weight associated with each of the clean faces. The images used for learning, from which the average image has been removed, are in fact a linear combination of clean faces.

$$\Phi\mathrm{i} = \sum_{i=l}^{l} Pidi \tag{3.9}$$

To find the associated weight, we make for each of the coordinates corresponding to a learning face.

$$Pi = d_i^T \Phi\mathrm{i} \tag{3.10}$$

This allows us to obtain for each of the M learning faces a vector πi, where is represent the itch face, and which informs us about the coefficient applied to each of the eigen faces.

Let's now move on to the task of recognizing a face of a subject. Once the image is taken, the image (column vector Γ) obtained is subtracted from the average image Ψ:

$$\Phi = \Gamma - \Psi \tag{3.11}$$

Puis nous trouvons les coordonnées de cette image dans l'espace réduit des faces propres :

$$\mathrm{Pi} = d_i^T \, \Phi\mathrm{i} \tag{3.12}$$

Which gives us in the end:

$$\Pi i = \begin{pmatrix} P1 \\ P2 \\ . \\ . \\ . \\ Pl \\ \\ \end{pmatrix}$$

We must now interpret the projection of the image to be analyzed in order to identify the subject. For this we will use a particular distance measurement, the distance of Mahalanobis. The advantage of this distance lies in the fact that it will give less weight to noisy components, and that it makes it possible to separate efficiently the axes for which the information can be better classified. It is defined by:

$$\mathrm{d(a,b)} = \sqrt{(a-b)^T Q^{-1} (a-b)} \tag{3.13}$$

Then, we compare the value of m found with a threshold value Δ, which should have been determined from testing on randomly chosen images, then comparing these values with the values obtained with learning faces, and determine if or not the analyzed image corresponds to a face present in the database. The choi

ce of this threshold depends on too many conditions (shooting images, level of precision desired for recognition, etc...

5. Applications of face detection

Facial detection technology is being used more frequently in photography as a way to help cameras autofocus on peoples' faces. One of the most prevalent uses of face detection is a facial recognition system, in order to establish identity. Facial recognition tools are used to secure phones and apps. They are also used by retail companies, airports, stadiums and other organizations in order to improve security. Also face detection is being used by some marketers in order to detect when people walk by a certain area. Face detecting systems can use algorithm to predict age, gender and other factors in order to serve up relevant advertisements.

6. Image fusion

The image fusion process is defined as gathering all the important information from multiple images and their inclusion into fewer images, usually a single one. This single image is more informative and accurate than any single source image, and it consists of all the necessary information. The purpose of image fusion is not only to reduce the amount of data but also to construct images that are more appropriate and understandable for the human and machine perception [DRA07] in computer vision, multisensory.

Image fusion is the process of combining relevant information from two or more images into a single image

6.1. Fusion levels

The image merge can be performed at any of the following three levels: signal level or pixel level, feature level, and decision level an elaboration on the first two levels) is presented in this chapter and the merge at decision level is always considered theoretical.

6.1.1. Pixel level fusion

In pixel-based fusion, source images are processed pixel by pixel. A merge ruler is applied to both pixels at the same location in the different source images. The resulting pixel value is then used to represent the pixel at that location in the composite image. The method is illustrated in figure 3.

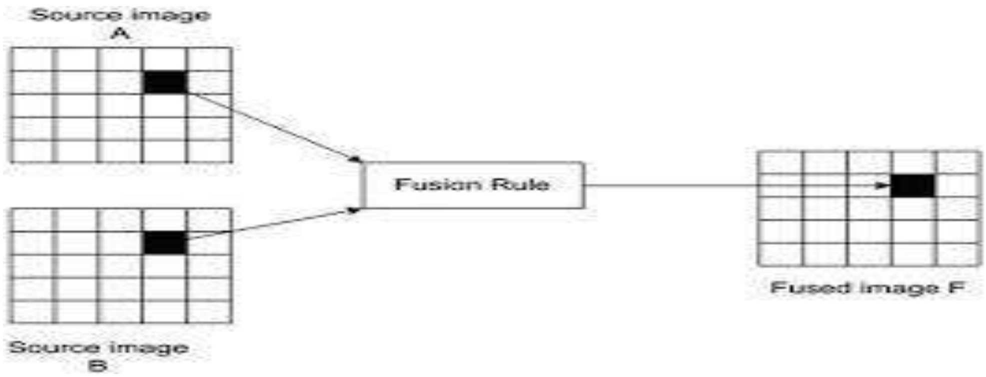


Figure 3.4. Image fusion at the pixel level

6.1.2 Feature-level fusion

The majority of application of a fusion scheme is interested in features within the image, not in the actual pixels. Therefore, it seems reasonable to incorporate feature information into process the fusion .

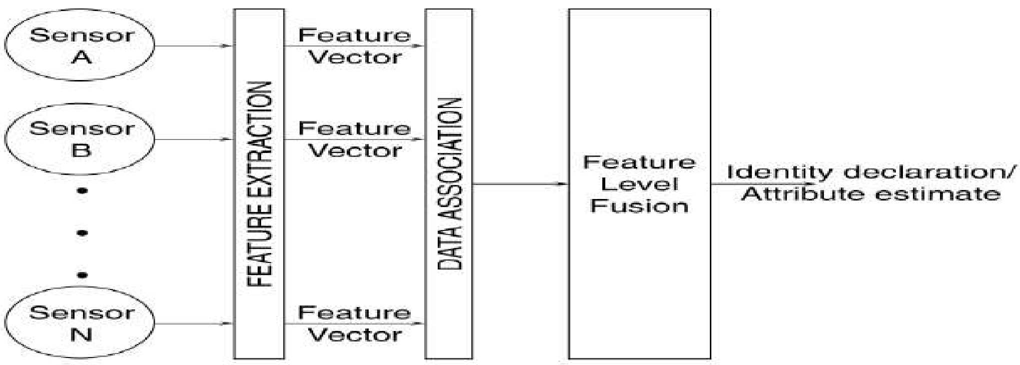


Figure 3.5.feature level fusion image .

6.1.3 Decision level fusion

Combines the result from multiple algorithms to yield a final fused decision. When the results from different algorithms are expressed as confidences rather than decisions, it is called soft fusion. Otherwise it is called hard fusion. Methods of decision fusion include voting methods, statistical methods and fuzzy logic based method [BAT11].

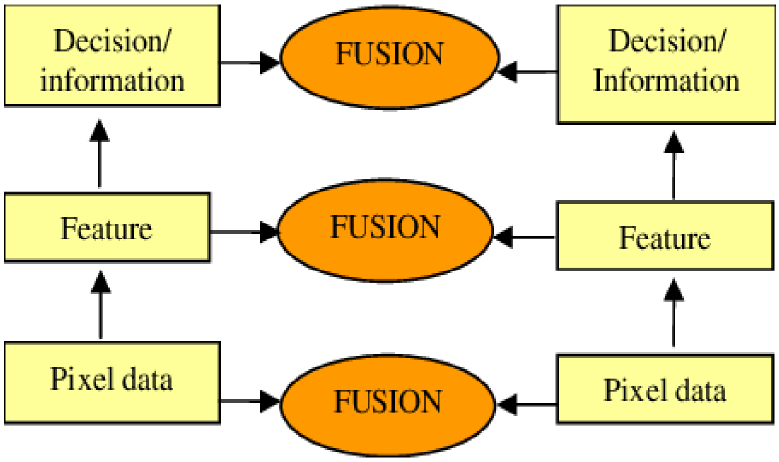


Figure 3.6. Pixel level, feature level and decision level fusion.

7. Methods fusion

There are many image fusion methods that can be used to produce high-resolution multispectral images from a high-resolution panchromatic image and low-resolution multispectral images

7.1. Weighted combination

The weighted combination is a simple and efficient calculation method. In addition, it removes the noise found in the source images due to the average function however, the fused sulfur image of the contrast and deletion part of the main characteristic.

$$F(x,y) = w_1.A(x,y) + w_2.B(x,y)$$

Where w1; w2 are weights of fused images.

7.2. The Discrete Cosine Transform (DCT)

The discrete cosine transform (DCT) helps separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality). The DCT is similar to the discrete Fourier transform: it transforms a signal or image from the spatial domain to the frequency domain.

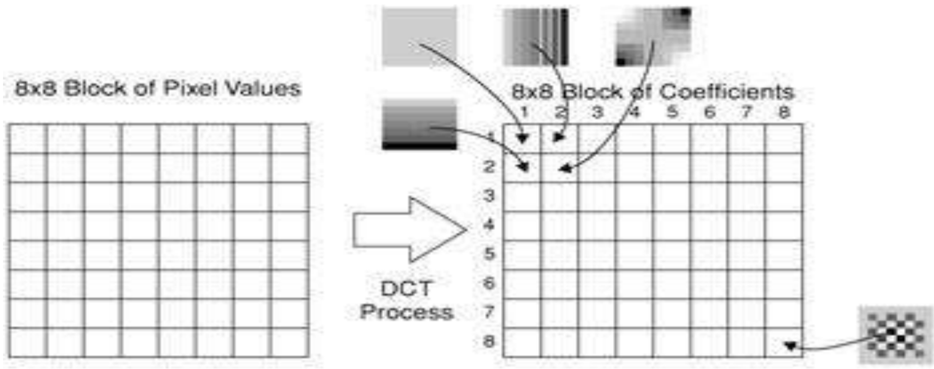


Figure 3.7. Method DCT.

The general equation for a 1D (N data items) DCT is defined by the following equation:

$$F(u) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} A(i).cos\left[\frac{\pi.u}{2.N}(2i+1)\right] f(i) \quad (3.14)$$

7.3. Pyramidal fusion technique

Pyramids are simply a convenient way to represent an image over a range of spatial resolutions. By combining the images at each level of the pyramid, the composite image, formed by pyramid reconstruction will have consistency over all resolutions.

When fusing two pyramids, each of the levels of the pyramids is fused into a composite level, resulting in a composite pyramid. Once the composite pyramid is formed, the: fused image of the source images is generated, employing the pyramid reconstruction techniques associated with the technique used to generate the source pyramids. For example, if pyramids A and B were generated from two source images, the composite image resulting from the fusion of pyramids A and B would be reconstructed from the composite pyramid C. Each level of the composite pyramid is defined as:

$$C_k = FUSE\ (A_k, B_k) \quad (3.15)$$

$$for\ K = n, n-1, n-2, \ldots\ 0, \quad (3.16)$$

And n is the number of levels in the pyramid FUSE is a function that converts the two images into the composite, using a fusion algorithm. The FUSE function was implemented two different ways; one way used a selection approach, while the other used a selection and averaging approach. The selection approach selected the pixel of highest contrast and it went into the composite image. The selection and averaging approach, called hybrid averaging and selection in this paper, selected a pixel when correlation was low, however, when correlation was high the pixels from the two source images were averaged for the composite pixel value[MEE99].

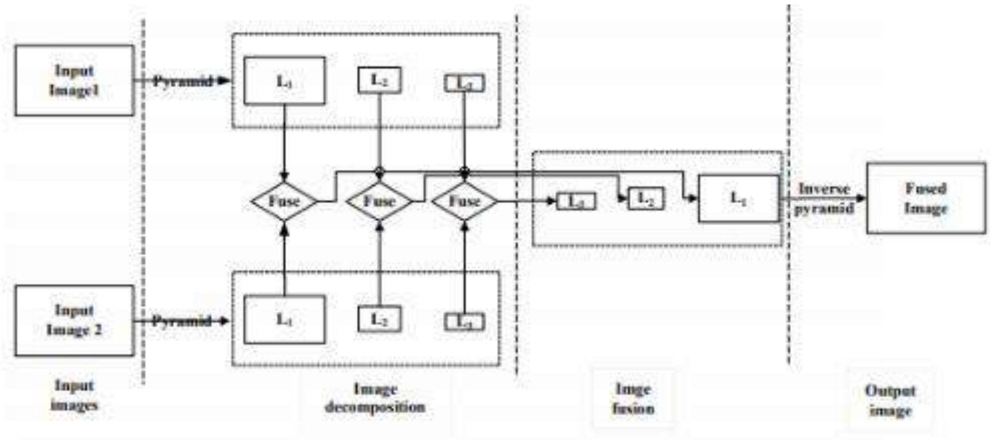


Figure 3.8. Method Pyramidal

8. Conclusion

In this chapter we tacked a overview about the face detection and image fusion, first we have see what's the face detection and some methods of, then we discussed the algorithm that we can use it to detect the faces and the application which use that.

Second the part of image fusion we resume the definition of image fusion and we talked about the levels of fusion and some methods that we will use it in the next chapter to fuse images with different depths.

Chapter 4

Results and discussion

1. Introduction

In this chapter, the implementation of our system is illustrated by showing the results and methods of changing the focus in the images to obtain several images with different depth. Next, two best-focused images will be combined using techniques (DCT, pyramidal, combination) these methods will be evaluated using some standard criteria. The obtained image applies an algorithm to detect the face in the image, as a final step for improvement.

2. Synoptic schema of our work

The structure of this work is suggested below in figure (4.1), which describes the general steps that are performed in the Matlab environment, for a solution to our problem which is to change the focus in the images and the face detection. The first step is to get a set of images with different depth. The second step is to merge the images some techniques are tested and then evaluated through several variables to choose the best ones. Finally, choose an algorithm face Detection ACP.

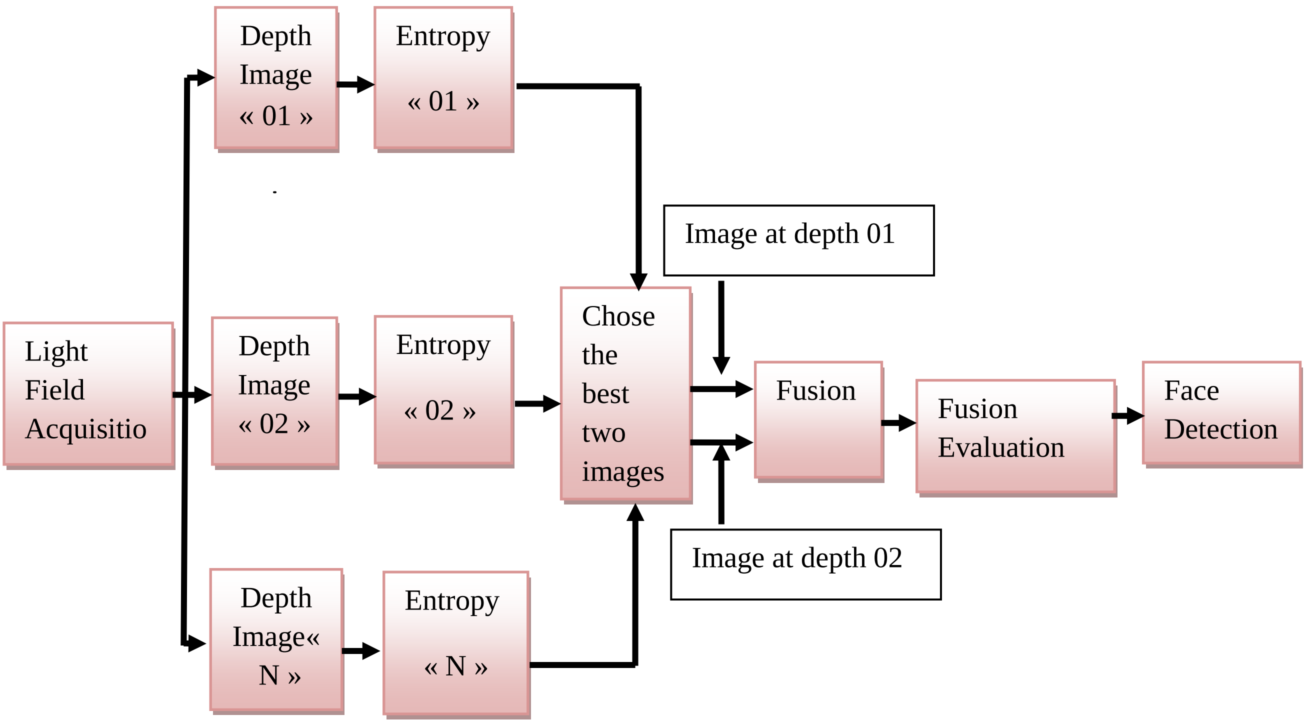


Figure 4.1. Synoptic schema.

3. Dataset

In this work we will use the database for three images taken by the plenoptic camera. This database contains a light field (LF) and camera parameters (x, y, z, LF, alpha, focal Length, V Max, U Max, lens, f Number,) [https://jpeg.org/plenodb/lf/epfl/].We choose three sets, the first one contains one face in defocus and the second contains two faces and the third contains 3 faces.

To get images from the light field we use the defocus function from the light field We get several images at different depths and we will choose the two best images to fuse by using three methods of fusion are DCT and combination and pyramidal we will see witch the good results of fusion are going to giving a good image with a best depth, finally it's come the step of- face detection.

4. Digital refocusing

4.1. Light field

The light field is a vector function that describes the amount of light flowing in every direction through every point in space. The direction of each ray is given by the 5D plenoptic function and the magnitude of each ray is given by the radiance.

The database for three images contains a light field (LF) and parameters, These parameters are represented en (x, y, z, (LF),alpha ,focal Length, V Max, U Max, lens ,f Number,) allow us to change the light field equation by changing the vial, which changes the focus of the image.

In this case, the light field has a dimension of equation (4.1).

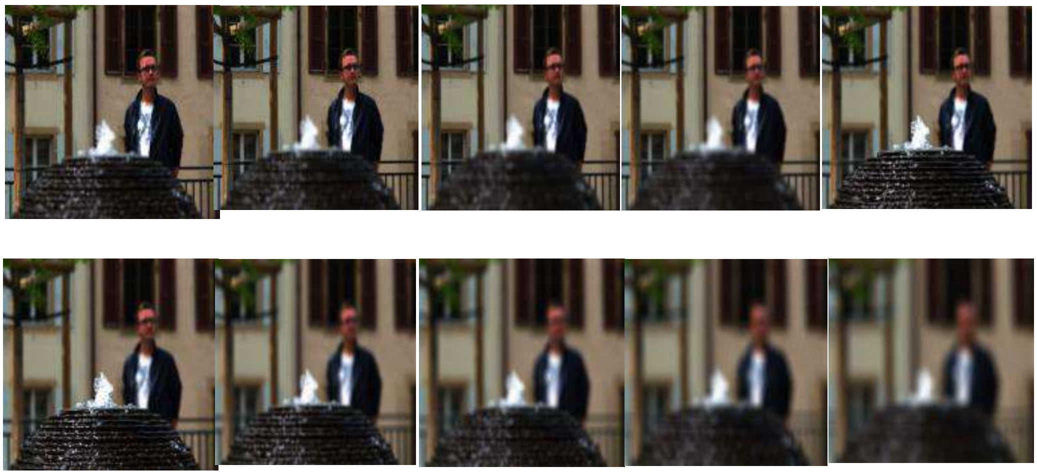
$$L(x, y, u, v, \lambda) \tag{4.1}$$

To get images from the light field we use the defocus function from the light field toolbox http://dgd.vision/Tools/LFToolbox/ [DAN13]. This toolbox for working with light field (plenoptic) imagery in Matlab. Features include decoding, calibration, rectification, color correction, basic filtering and visualization of light field images. New in version 0.4 are some linear depth/focus and denoising filters.

The digital defocusing function is based on the following equation (4.2):

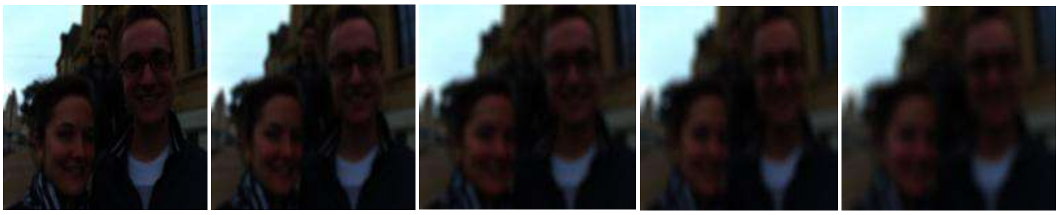
$$E_{(\alpha,F)}(x',y')=\frac{1}{\alpha^2F^2}\iint L_F\left(u\left(1-\frac{1}{\alpha}\right)+\frac{x'}{\alpha'}v\left(1-\frac{1}{\alpha}\right)+\frac{y'}{\alpha},u,v\right)dudv \qquad (4.2)$$

Results of this part: Many images at different depths: -2/9,-4/9,-8/9,-12/9,-16/9,-20/9, +2/9, +4/9, +6/9, +8/9.



(a) Set 01

(b) Set 02



(c) Set 03

Figure 4.2.Images at different depths" (a) Set 01,(b) Set 02,(c) Set 03".

4.2. Selecting images

Choosing the best images to fuse:

Considering"n"depth images obtained from one single capture, represented by $\{ I_{d1}, I_{d2}, \ldots, I_{dk}, \ldots, I_{dn} \}$, The best focused image is obtained such that the best focus corresponds to the maximum energy .based on computed energy for each depth image according to equation (4.3) (4.4), two images corresponding to highest energy and second highest energy are second highest energy and second highest energy are selected

$$E_{d_{sel1}} = \max\{E_{1_{Total}}, E_{2_{Total}} \ldots\ldots E_{n_{Total}}\} \tag{4.3}$$

$$E_{d_{sel2}} = \max\{E_{1_{Total}}, E_{2_{Total}} \ldots\ldots E_{n_{Total}}\} \tag{4.4}$$

Such that:

The images corresponding to$E_{d_{sel\,1}}$and $E_{d_{sel\,2}}$are used to obtain fused image with higher amount of information [RAJ15].

4.3. Entropy (EN)

The entropy been used by Leung et al. (2001) to measure the performance of image fusion. The entropy of an image is given in equation (4.5):

$$En = \sum_{i=0}^{G} \left(\mathrm{P(i)Log(P(i)}\right) \tag{4.5}$$

Where: G is the number of gray levels in the image and $(\mathrm{P(i)})$ is the normalized probability of occurrence of each gray level.

5. Image Fusion

More recently, various new fusion techniques were proposed for fusing images of different focal length [RAG13]. In this work, we perform an empirical evaluation of such fusion techniques for a face detection system using two best focused images from the set of depth images. Furthermore, we also provide the verification performance benchmark with well-established all in focus images [RAG13].

5.1. Visual Results

The resultants of image fusion using DCT, Pyramid fusion and combination. The two sets image reference of this work is provided from plenoptic camera the results of fusion algorithms are presented in figures 4.2, 4.3 and 4.4.

(a) Image at depths 01 (b) Image at depth 02 (c) Combination fusion result

(a) Image at depth 01 (b) Image at depth 02 (c) Pyramidal fusion result

(a) Image at depth 01 (b) Image at depth 02 (c) DCT fusion result

Figure 4.3. Results of fusion for "01 people".

(a) Image at depth 01 (b) Image at depth 02 (c) Combination fusion result

(a) Image at depth 01 (b) Image at depth 02 (c) Pyramidal fusion result

(a) Image at depth 01 (b) Image at depth 02 (c) DCT fusion result

Figure 4.4. Results of fusion for "02 people"

(a) Image at depth 01 (b) Image at depth 02 (c) Combination fusion result

(a) Image at depth 01 (b) Image at depth 02 (c) Pyramidal fusion result

(a) Image at depth 01 (b) Image at depth 01 (c) DCT fusion result

Figure 4.5. Results of fusion for “03 people”.

5.2. Commentary about the fusion schemas

In this case, we can see that the good results of fusion are going to DCT methode ,which are giving a good image brightness with nice detail and sharpness which we can see that in figure 4.3 “01 people” which can we see the person and the fountain more clearly and also in figure 4.4”02 People” Where you can see the first person with the first focus and the second person with the second focus clearly, combined are giving a high quality image and detail which is given a results are very clear and have a good brightness with high quality resolution. In the other hand combination are giving a high quality image and detail which is closely to the results of DCT. But, pyramidal is not helpful to us in this work because the results in some figures are not clear exactly in figure 4.3“01 people”.

6. The performance analysis of fusion algorithms

The multi-sensor image fusion methods are directed towards applications in the areas of surveillance and navigation. In these applications, due to the real-time nature of the scene being imaged, there are no ground truth data available.

In order to analyze and evaluate the fusion schemes, objective measure techniques should be introduced and used. Below; a total of 5 fusion evaluation metrics are introduced and explained thoroughly in theory:

6.1 Mutuel Information (MI)

It is calculated by defining the joint histogram of the source imageI_{ir}, $I_{vis}$and the fused image I_{fuse} as $P(fuse, ir)$ and $P(fuse, vis)$.The mutual information between the source image. and the fused image is given the equation (4.6) and (4.7):

$$\text{Mi1}(\text{fuse}, \text{ir}) = -\sum P(fuse, ir) Log \left\{\frac{P(fuse,ir)}{P(fuse).P(ir)}\right\} \tag{4.6}$$

$$\text{Mi2}(\text{fuse}, \text{vis}) = -\sum P(fuse, vis) Log \left\{\frac{P(fuse,vis)}{P(fuse).P(vis)}\right\} \tag{4.7}$$

Where: $P(fuse, ir)$ and $P(fuse, vis)$are the joint histograms of the source image I_{ir} , I_{vis} and the fused imageI_{fuse}. The fusion algorithm efficiency is determined by the metric MI which is defined by (4.8):

$$M_{i=}M_{i_1}(fuse, ir) + M_{i_2}(fuse, vis) \tag{4.8}$$

6.2 Power Signal-to-Noise Ratio (PSNR)

Maruthi and Suresh in (2007) have published results using the Non-reference PSNR defined as the ratio of the mean pixel value to the standard deviation of the pixels:

$$\text{PSNR} = \frac{Mean\ Pixel\ Intensity}{Standard\ Deviation} \tag{4.9}$$

This metric has more or less the same value as the standard deviation and is not efficient in discriminating the quality of an image.

6.3. RMSE

To simplify, we assume that we already have n samples of model errors ϵ calculated as (e_i, i = 1, 2... n). The uncertainties brought in by observation errors or the method used to compare model and observations are not considered here. We also assume the error sample set ϵ is unbiased. The RMSE are calculated for the data set as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^1} \tag{4.10}$$

The underlying assumption when presenting the RMSE is that the errors are unbiased and follow a normal distribution.

7. Empirical Results

Below, tables which summarize the results of the entire objective evaluation criterion applied on all dataset considered image set with 3 different image fusion methods.

Table 4.1. Presents the results of evolutions metrics and averages values of deferent fusion algorithms proposed for the **“01 People”**.

Fusion methode	EI	MI2	PSNR	SF	RMSE
Combination	0.1508	2.5269	33.3470	0.0334	0.0214
Pyramid	0.1556	2.7617	**30.5200**	0.0387	0.0297
DCT	**3.1526**	**5.0392**	27.9674	**6.3258**	**10.1500**

Table 4.2. Presents the results of evolutions metrics and averages values of deferent fusion algorithms proposed for the **“02 People”**.

Fusion methode	EI	MI	PSNR	SF	RMSE
Combination	0.0755	**3.2846**	**37.9568**	0.0180	0.0126
Pyramid	0.0902	3.1477	32.1696	0.0233	0.0245
DCT	**25.6431**	2.4833	29.9478	**8.4362**	**8.0806**

Table 4.3. Presents the results of evolutions metrics and averages values of deferent fusion algorithms proposed for the **"03 People"**.

Fusion methode	EI	MI	PSNR	SF	RMSE 2
Combination	0.0540	3.2183	41.7000	0.0173	0.0082
Pyramid	0.0539	**3.2983**	**44.3879**	0.0161	0.0060
DCT	**8.5694**	1.8696	28.9166	**9.0349**	**9.0992**

7.1. Commentary about the table

From the table we choose this method since it has a higher value of PSNR and EI especially in **"03 People"** and **"04 People"** and smaller than MI But, generally DCT is the great fusion schema which gives a good resolution and brightness with a high quality image than combination , so it very helpful and suitable for our work

8. Results of Face Detection

This is the final step in engineering this work. We used the algorithm described in Chapter 03. The ACP, PCA (Principal Component Analysis) algorithm. This algorithm relies on well-known statistical properties and uses the linear algebra it is relatively.

The main idea is to express the M training images according to a base of particular orthogonal vectors, containing information independent of a vector to another. These new data are thus expressed in a more appropriate way to the face, the face method, which is used in the technical analysis of the main component analysis, which is based on the peculiarities of the analyzed method. , since the whole face is then analyzed. Our technical recognition will be used in the main component analysis method.

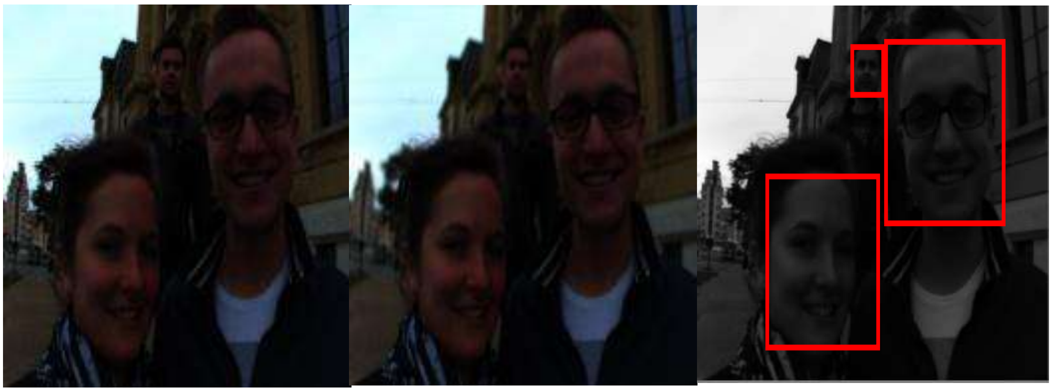
The results are given below in figure 4.5, 4.6 and figure 4.7.

(a) Image at depth 01 (b) Image at depth 01 (c) Face detection results

Figure 4.5. Results of face detection**"01 People".**

(a) Image at depth 01 (b) Image at depth 01 (c) Face detection results

Figure 4.6. Results of face detection **"02 People"**



(a) Image at depth 01 (b) Image at depth 02 (c) Face detection results

Figure 4.7. Results of face detection **"03 People".**

8.1. Commentary about the results

With these resulting images we can finally say. The problem of out-of-focus photography has been solved, we provide the results obtained by employing a face detection system frame, all the results are good and give us some missing information in the source image, and we can improve the image quality with this method

9. Conclusion

The problem of out-of-focus imaging has been solved using light field cameras, and this has been well adopted in the domain of biometrics for face detection .In this work, we aim to employ a method to select the images when more than two images corresponding to different focus are present. Then, a fusion technique will be used to fuse the chosen images. The resulted image will contain all faces from different focuses presented in the same image, which solves the problem of face recognition from out-of-focus imaging.

General conclusion

The light field is unlimited technologies and this is just its beginning in the imaging that make us be so excited for the coming challenge in this technologies .In this work we tried to simplify the concept of the light field which makes understood to everyone, we started in first chapter with the conventional imaging and its problems , those problems whose makes this challenge to discover the light field imaging , than we saw the light filed imaging meaning , second the plenoptic function as defined by adelson and bergen, We represented also the light field parameterization, as we saw that we choose the two planes parameterization because of its simplicity, we saw also an overview to the light field acquisition and its application

Chapter two displayed the image formation and digital refocusing, the image formation and plenoptic camera records the light field and the imaging equations and also digital refocusing with some examples. Finally the image synthesis algorithms are presented.

The chapter three shows the image fusion and face detection, then we discussed the face detection algorithm, the part of image fusion talked about the definition of image fusion and the levels of fusion and some methods that we used it in the next chapter to fuse images with different depths .

The chapter four is a results and discussion of this work , as we say before that we chose about 30 images with a different depths contained more than one face, one is clear and the other is not, than we applied three methods of fusion are DCT and combination and pyramidal to fuse the best two images , we knew those two by the result of the methods , the final result showed as a clear image with all the faces , for the evaluation we used the parameters : PSNR (Power Signal-to-Noise Ratio),RMSE,MI(mutual information) ,EI(edge intensity),SF(space frequency) ,and after this evaluation it was clear that the best method to fuse the set of images is DCT .

Scientists are trying to make future cameras based on these principles will be physically simpler, capture light more quickly, and provide greater flexibility in finishing photographs. There is a lot of work to be done on re-thinking existing camera components in light of these new capabilities. With larger-aperture lenses, it may be possible to use a weaker flash system or do away with it in certain scenarios. Similarly, the design of the auto-focus system will change in light of digital refocusing and the shift in optimal lens focus required by selectable refocusing power.

Perhaps the greatest upheaval will be in the design of the photosensor. We need to maximize resolution with good noise characteristics not an easy task. And the electronics will need to read it out at reasonable rates and store it compactly. This is the main price behind this new kind of imaging: recording and processing a lot more data. Fortunately, these kinds of challenges map very well to the exponential growth in our capabilities for electronic storage and computing power.

Bibliography

[ADE91] ADELSON, E. H. «The Plenoptic Function and the Elements of Early Vision", Computational Models of Visual. *Processing, Chap. 1, Edited by M. Landy and JA Movshon*, 1991.

[ADE92] ADELSON, Edward H. ET WANG, John YA. Single lens stereo with a plenoptic camera. *IEEE transactions on pattern analysis and machine intelligence*, 1992, vol. 14, no 2, p. 99-10

[AGR10]AGRAWAL, Amit, VEERARAGHAVAN, Ashok, et RASKAR, Ramesh. Reinterpretable Imager: Towards Variable Post-Capture Space, Angle and Time Resolution in Photography. In: *Computer Graphics Forum*. Blackwell Publishing Ltd, 2010. p. 763-772.

[BIC15] BICHAL, Abhishek. *Development of 3D background oriented schlieren with a plenoptic camera*. 2015. Thèse de doctorat.

[BOL14] BOLAN, Jeffrey, JOHNSON, Kyle C., et THUROW, Brian S. Preliminary investigation of three-dimensional flame measurements with a plenoptic camera. In: *30th AIAA Aerodynamic Measurement Technology and Ground Testing Conference*. 2014. p. 2520.

[GOR96] GORTLER, Steven J., GRZESZCZUK, Radek, SZELISKI, Richard, *et al.* The lumigraph. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996. p. 43-54.

[HAN06] HANRAHAN, Pat et NG, Ren. Digital correction of lens aberrations in light field photography. In: *International Optical Design Conference*. Optical Society of America, 2006. p. WB2.

[IHR 16] IHRKE, Ivo, RESTREPO, John, et MIGNARD-DEBISE, Lois. Principles of Light Field Imaging: Briefly revisiting 25 years of research. *IEEE Signal Processing Magazine*, 2016, vol. 33, no 5, p. 59-69.

[IHR10] IHRKE, Ivo, WETZSTEIN, Gordon, et HEIDRICH, Wolfgang. A theory of plenoptic multiplexing. In : *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010. p. 483-490.

[LEV 96] LEVOY, Marc et HANRAHAN, Pat. Light field rendering. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996. p. 31-42.

[LEV 06] LEVOY, Marc. Light fields and computational imaging. *Computer*, 2006, vol. 39, no 8, p. 46-55.

[MEE99] MEEK, T. R. An Analysis of Pyramidal Image Fusion Techniques. 1999.

[RAG13] RAGHAVENDRA, Remarchandera, RAJA, Kiran B., YANG, Bian, *et al.* A novel image fusion scheme for robust multiple faces recognition with light-field camera. In: *Information Fusion (FUSION), 2013 16th International Conference on*. IEEE, 2013. p. 722-729.

[RAG13] [RAGRAGHAVENDRA, Remarchandera, YANG, Bian, RAJA, Kiran B., *et al.* A new perspective—Face recognition with light-field camera. In: *Biometrics (ICB), 2013 International Conference on*. IEEE, 2013. p. 1-8.

[TOL06] TOLBA, A. S., EL-BAZ, A. H., et EL-HARBY, A. A. Face recognition: A literature review. *International Journal of Signal Processing*, 2006, vol. 2, no 2, p. 88-103.

[TUR91] TURK, Matthew A. et PENTLAND, Alex P. Face recognition using eigenfaces. In: *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. IEEE, 1991. p. 586-591.

[TUR91] TURK, Matthew et PENTLAND, Alex. Eigen faces for recognition. *Journal of cognitive neuroscience*, 1991, vol. 3, no 1, p. 71-86.¨

[UNG03] UNGER, Jonas, WENGER, Andreas, HAWKINS, Tim, *et al. Capturing and rendering with incident light fields*. UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY CA INST FOR CREATIVE TECHNOLOGIES, 2003.

[WET11] WETZSTEIN, Gordon, IHRKE, Ivo, LANMAN, Douglas, *et al.* Computational plenoptic imaging. In: *Computer Graphics Forum*. Blackwell Publishing Ltd, 2011. p. 2397-2426.

[WIL01]WILBURN, Bennett S., SMULSKI, Michal, LEE, Hsiao-Heng Kelin, *et al.* Light field video camera. In: *Media Processors 2002*. International Society for Optics and Photonics, 2001. p. 29-37.

[YAN02] YANG, Jason C., EVERETT, Matthew, BUEHLER, Chris, *et al.* A real-time distributed light field camera. *Rendering Techniques*, 2002, vol. 2002, p. 77-86.

Abstract

Light field imaging has emerged as a technology allowing to capture richer visual information from our world. Unlike conventional cameras, which record the 2D projection of the light rays by angularly integrating the rays at each pixel, a light field cameras capture not only the light intensity in the image plane but also the information about the direction of light rays. Although face detection is a widely accepted form of biometrics due to its ease of capture, it still suffers from out-of-focus imaging due to the limited depth-of-field found in conventional 2D imaging systems. In this work, and in order to overcome such a challenge, we use the light field to obtain a number of depth images. Then, a fusion technique is used to fuse the best focus images. The resulted image contains all faces from different focuses presented in the same image, which solves the problem of face detection from out-of-focus imaging

Résumé

L'imagerie de champ lumineux est apparue comme une technologie permettant de capturer des informations visuelles plus riches de notre monde. Contrairement aux caméras conventionnelles, qui enregistrent la projection 2D des rayons lumineux en intégrant angulairement les rayons à chaque pixel, les caméras à champ lumineux capturent non seulement l'intensité lumineuse dans le plan image mais aussi l'information sur la direction des rayons lumineux. Bien que la détection de visage soit une forme largement acceptée de biométrie en raison de sa facilité de capture, elle souffre toujours d'une imagerie floue en raison de la profondeur de champ limitée des systèmes d'imagerie 2D conventionnels. Dans ce travail, et afin de surmonter un tel défi, nous utilisons le champ lumineux pour obtenir un certain nombre d'images de profondeur. Ensuite, une technique de fusion est utilisée pour fusionner les meilleures images de mise au point. L'image obtenue contient toutes les faces de différents foyers présentés dans la même image, ce qui résout le problème de la détection de visage à partir de l'imagerie non focalisée.

الملخص

لقد ظهر التصوير في مجال الضوء كإحدى التقنيات التي تتيح التقاط معلومات بصرية أكثر ثراءً من عالمنا. على عكس الكاميرات التقليدية ، التي تسجل إسقاط بعدين فقط من أشعة الضوء عن طريق دمج أشعة الشمس في كل بكسل ، تلتقط الكاميرات الحقلية الضوء ليس فقط كثافة الضوء في مستوي الصورة ولكن أيضا معلومات حول اتجاه الأشعة الضوئية. على الرغم من أن اكتشاف الوجه هو شكل مقبول على نطاق واسع من القياسات الحيوية نظرًا لسهولة التقاطه ، فإنه لا يزال يعاني من التصوير خارج التركيز نظرًا لحدود العمق المحدودة الموجودة في أنظمة التصوير التقليدية ثنائية الأبعاد. في هذا العمل ، وللتغلب على هذا التحدي ، نستخدم حقل الضوء للحصول على عدد من صور العمق. ثم ، يتم استخدام تقنية الانصهار لدمج أفضل صور التركيز. تحتوي الصورة الناتجة على جميع الوجوه من بؤر مختلفة معروضة في الصورة نفسها ، والتي تحل مشكلة اكتشاف الوجوه من التصوير خارج التركيز