The Road to Immersive Communication

J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, “The road to immersive communication,” Proceedings of the IEEE, vol.100, no.4, pp.974-990, 2012.

Outline

1. Introduction
2. History
3. Sensory signal processing
   a. Visual
   b. Auditory
   c. Other senses(haptics, smell, taste)
4. Communication aspects
5. User exprience(QoE)

Outline

1. **Introduction**
2. History
3. Sensory signal processing
    a. Visual
    b. Auditory
    c. Other senses(haptics, smell, taste)
4. Communication aspects
5. User exprience(QoE)

Introduction

Immersive communication means exchanging natural social signals with remote people, as in face-to-face meetings, or experiencing remote locations in ways that suspend disbelief in being there.

The quality of an immersive communication system is judged by its impact on the humans who use it(i.e., exceedingly difficult to develop ojective metrics).

Visual, audio, haptics, smell, and taste.

Introduction

Focus on visual and audio, including capture, render, and process.

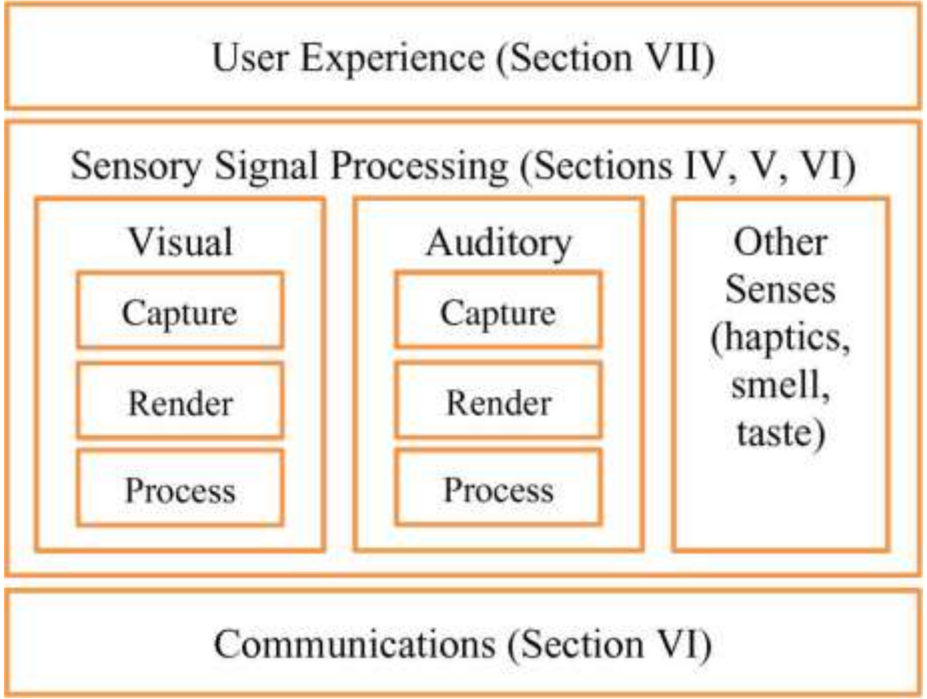

User Experience (Section VII)
Sensory Signal Processing (Sections IV, V, VI)
Visual
Capture
Render
Process
Auditory
Capture
Render
Process
Other Senses (haptics, smell, taste)
Communications (Section VI)

Fig. 1. The elements of an immersive communication system.

Outline

1. Introduction
2. **History**
3. Sensory signal processing
    a. Visual
    b. Auditory
    c. Other senses(haptics, smell, taste)
4. Communication aspects
5. User exprience(QoE)

History

Anticipates immersive communication in this 1878 cartoon.

Telephone(1876 or early)

Television broacast, Berlin Olympic Games(

AT&T Picturephone(1964)

PC conference, Microsoft(1995)

Skype(2006)

To be continued.

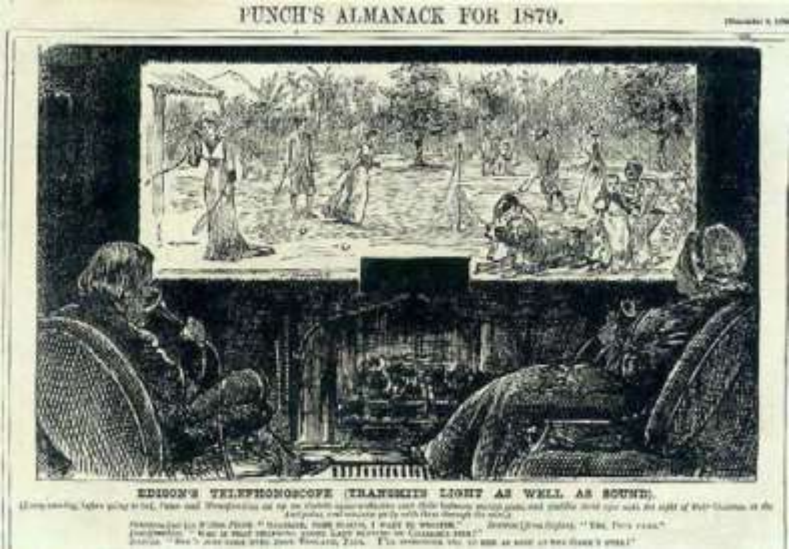

PUNCH'S ALMANACK FOR 1879.
EDISON'S TELEPHONOSCOPE (TRANSMITS LIGHT AS WELL AS SOUND).

Fig. 2. George du Maurier anticipates immersive communication in this 1878 cartoon.

Framework

Communcation system, share voice and life-size video(feeling of being same room).

Collaboration system, share information such as emails, docs and applications.

Combine system(shared environment for collaboration)

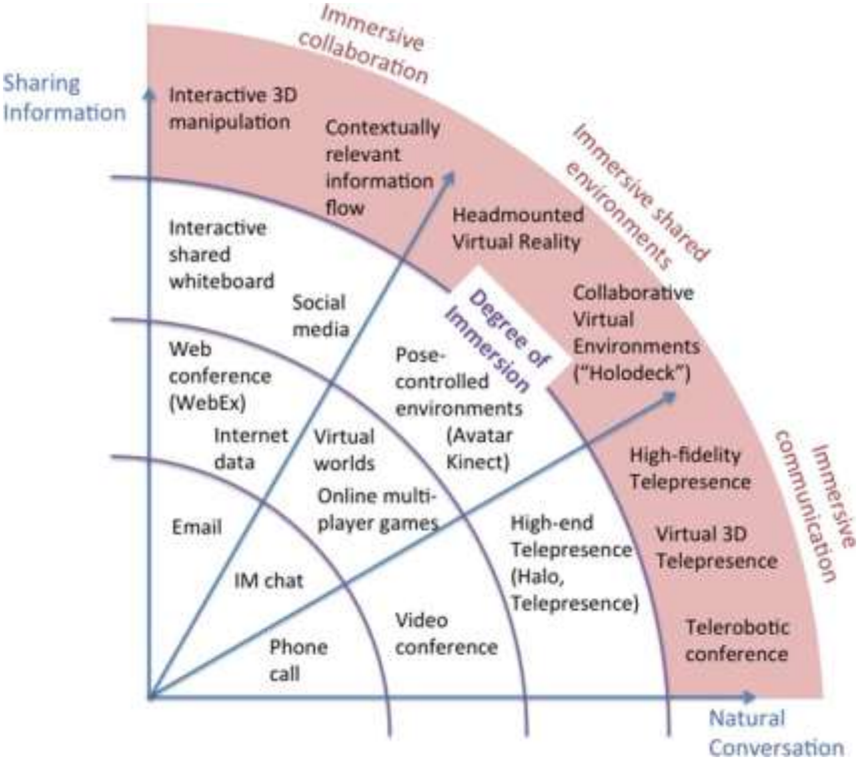

Immersive collaboration
Sharing Information
Interactive 3D manipulation
Contextually relevant information flow
Immersive shared environments
Headmounted Virtual Reality
Interactive shared whiteboard
Social media
Degree of Immersion
Collaborative Virtual Environments ("Holodeck")
Web conference (WebEx)
Pose-controlled environments (Avatar Kinect)
Internet data
Virtual worlds
Immersive communication
High-fidelity Telepresence
Online multi-player games
Email
High-end Telepresence (Halo, Telepresence)
Virtual 3D Telepresence
IM chat
Video conference
Phone call
Telerobotic conference
Natural Conversation

Fig. 3. Communication and collaboration framework.

Outline

1. Introduction
2. History
3. **Sensory signal processing**
    a. **Visual**
    b. Auditory
    c. Other senses(haptics, smell, taste)
4. Communication aspects
5. User exprience(QoE)

Visual Rendering

To provide an immersive visual experience for each viewer.

Key attributes:

each eye of each viewer sees an appropriate view(stereoscopy)

view changes as the viewer moves(motion parallax)

objects move in and out of focus as the viewer changes his/her focal plane

Visual Rendering

Flat display:

Light-field dispaly

Conventional display(RGB)

Stereoscopic display:

Stereoscopic 3-D display(two-view)

Autostereoscopic

Curved display

Head-mounted display

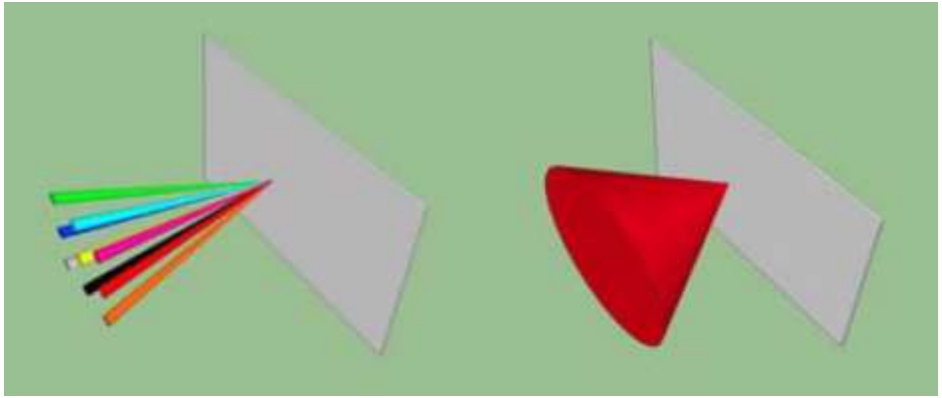


Fig. 4. For each spatial location in a rectangular display, light-field(left) and conventional display(right).

Visual Capture

Light-field camera

Places a 2-D lens array in front of a 2-D sensor.

Each lens then focuses different angles onto different sensors in the 2-Dsensor, thereby capturing both spatial and angular information.

4-D plenoptic function, P(x, y, θ, Ø), location(x, y) and direction(θ, Ø)

Conventional camera

Capture only a 2-D projection of the 4-D plenoptic function.

A stereo camera incorporates two lens and two 2-D sensors.

Visual Interaction

Face-to-face discussions employ important nonverbal signals such as eye contact, gaze direction, and gestures such as pointing.

Challenges:

- See-through display
- Robot
- Remote control camera
- View synthesis algorithms

Outline

1. Introduction
2. History
3. **Sensory signal processing**
    a. Visual
    b. **Auditory**
    c. Other senses(haptics, smell, taste)
4. Communication aspects
5. User exprience(QoE)

Audio Rendering

An audio experience is created by establishing a sound pressure field in each ear canal.

Psychoacoustics plays a key role in current approaches; human insensitivity to certain acoustic details.

Headset:

- Real-time rendering of spatially consistent binaural audio for headsets has proved difficult
- Synthesize sound

Audio Rendering

Surrounding loudspeakers:

- Calculate transfer function $(NM)^{-1}$ that the N loudspeaker excitation signals needed to deliver the desired sound signals to the M target eardrums.

Challenges:

- The transfer functions are difficult to determine with sufficient accuracy.
- The transfer functions can change quickly due to head movements.
- Inversion can be ill-conditioned.
- The signal processing demands a low-latency implementation.

Audio Capture

Capture of sound field elements that are relevant for remote rendering.

Immersive rendering systems that employ loudspeakers must solve two coupled difficult problems:

- How to induce desired sounds in the ears of the listeners
- How to remove the resulting contamination from the microphone signals.

By case:

- Capture methods isolates clean versions of individual sound sources.
- Captures a sound field at a location or over an area

Audio/Video Correspondence

1-1 meeting:

Video is powerful cue for audio, particularly when the video is perceived as relevant to the audio.

Thus it would seem that one could relax some constraints on audio rendering when video is present.

Party:

Audio should be good enough in this situation to allow the cocktail party effect to take place.

Depend on the context and purpose of the communication.

Outline

1. Introduction
2. History
3. **Sensory signal processing**
    a. Visual
    b. Auditory
    c. **Other senses(haptics, smell, taste)**
4. Communication aspects
5. User exprience(QoE)

Other senses

Haptics/touch

Weber’s law

Just noticeable difference(JND)

Changes less than the JND would not be perceived and hence do not need to be transmitted.

Smell

Taste

Outline

1. Introduction
2. History
3. Sensory signal processing
    a. Visual
    b. Auditory
    c. Other senses(haptics, smell, taste)
4. **Communication aspects**
5. User exprience(QoE)

Communications

New standard, including better compression algorithm and transport protocol.

Compute capablility(faster, cloud computing)

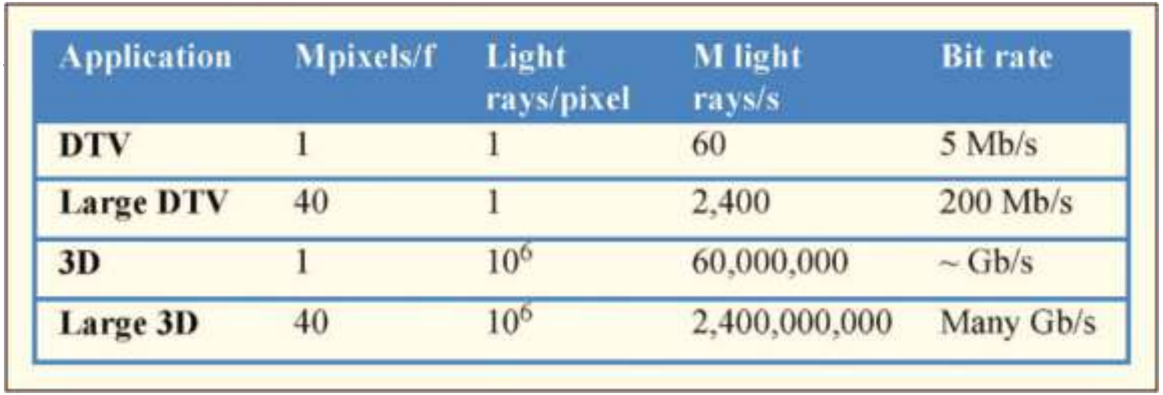
Powerful graphics proces

Parallel computing

Storage

Network bandwidth

New sensors



Application	Mpixels/f	Light rays/pixel	M light rays/s	Bit rate
DTV	1	1	60	5 Mb/s
Large DTV	40	1	2,400	200 Mb/s
3D	1	10^6	60,000,000	~ Gb/s
Large 3D	40	10^6	2,400,000,000	Many Gb/s

Table. 1. Example of the Large Increase in Raw and Compressed Data Rates for Going From a Conventional Digital TV Video (Row 1) to a Large Display Providing an Immersive Experience (Row 2), and to Future 3-D Light Field Displays (Rows 3 and 4).

Outline

1. Introduction
2. History
3. Sensory signal processing
    a. Visual
    b. Auditory
    c. Other senses(haptics, smell, taste)
4. Communication aspects
5. **User exprience(QoE)**

User experience

As mentioned in the **Introduction**.

To achieving a high QoE is inducing in each participant an intended illusion or mental state.

Performance level: eye contact, gaze awareness, facial expression and so on.

Psychometric level: feel, environment, surroundings.

Psychophysical level: Weber’s law, JND.

Outline

1. Introduction
2. History
3. Sensory signal processing
    a. Visual
    b. Auditory
    c. Other senses(haptics, smell, taste)
4. Communication aspects
5. User exprience(QoE)

Conclusion

Humans are social animals, and have evolved to interact with each other most effectively face-to-face.

Colocation enables us to exhibit and interpret various nonverbal signals and cues.

Global economic, to reduce environmental impact, to reduce travel costs and fatigue, and to richer collaboration in business.

Technical trends are rapidly accelerating progress towards new and more immersive communication systems.

Outline

1. Introduction
2. History
3. Sensory signal processing
    a. Visual
    b. Auditory
    c. Other senses(haptics, smell, taste)
4. Communication aspects
5. User exprience(QoE)

References

1. J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, “The road to immersive communication,” Proceedings of the IEEE, vol.100, no.4, pp.974-990, 2012

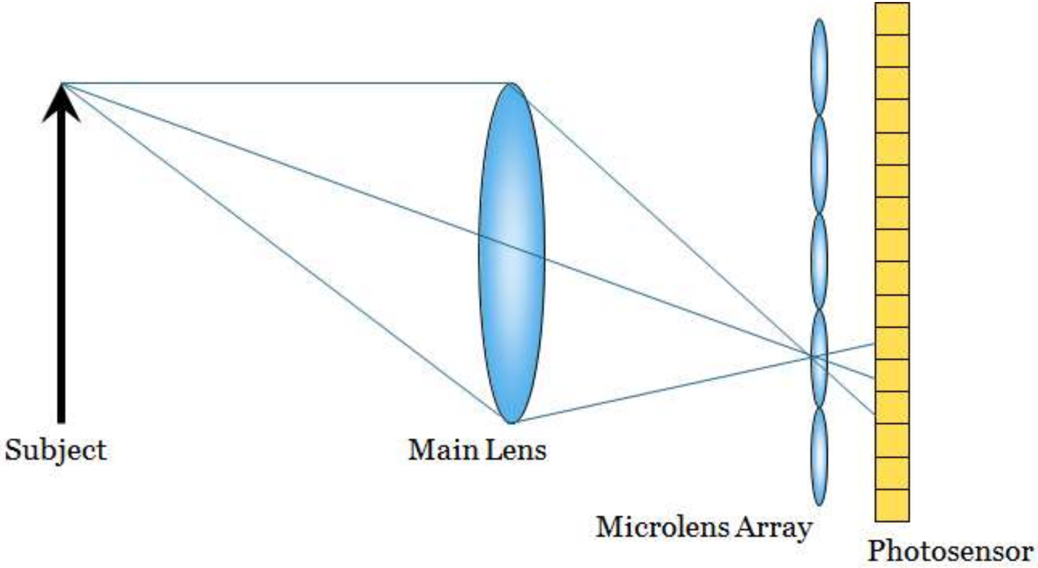
1. R. Ng, “Light Field Photography,” http://graphics.stanford.edu/papers/lfcamera/, 2005.

Backup

Light-field

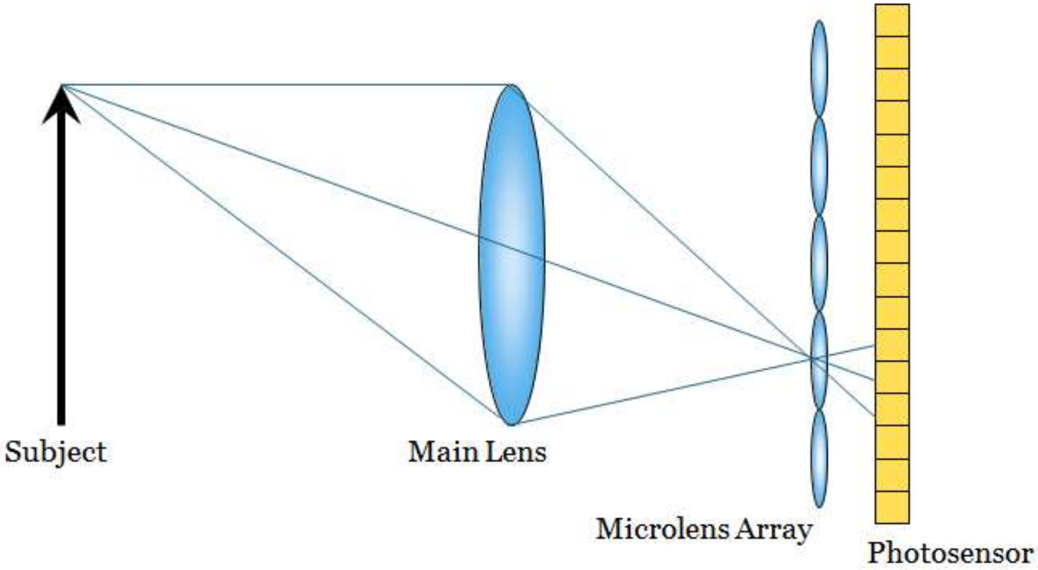
Rays of lights form a single point on the subject are brought to a single convergence point on the focal plane of the microlens array.

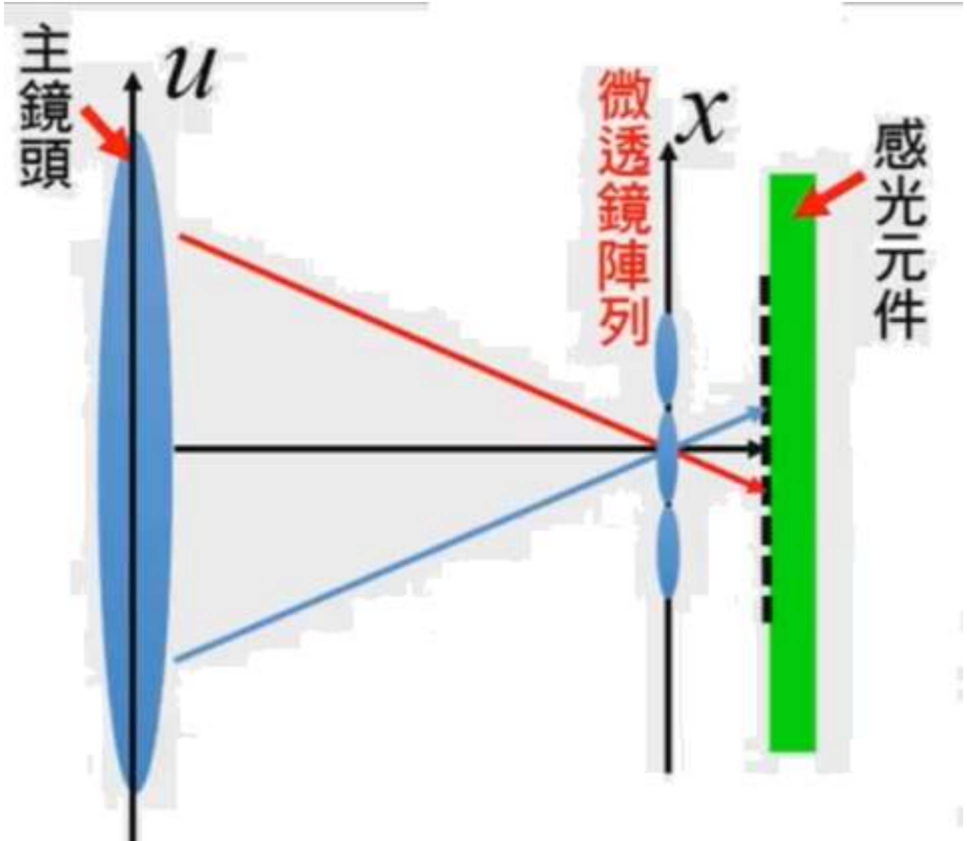
The microlens at the location separates these rays of light based on its direction.

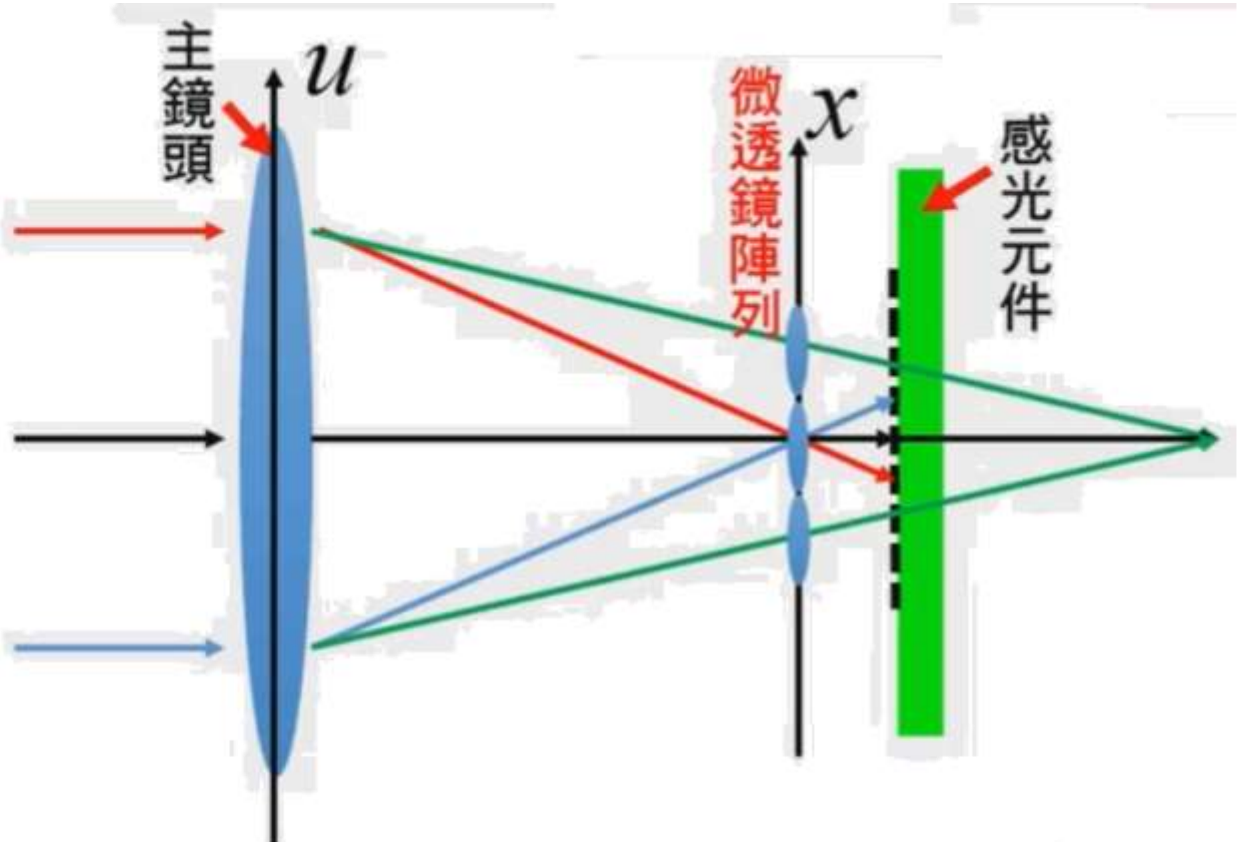


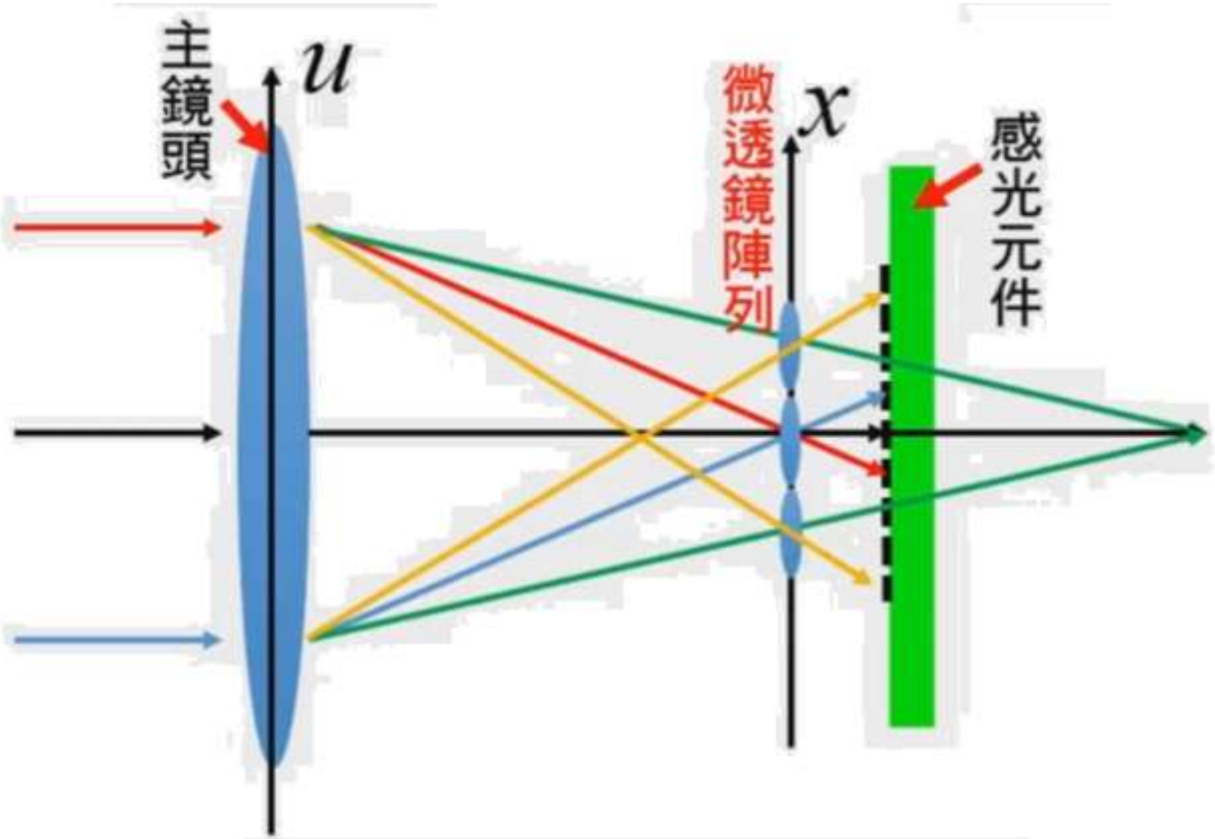
Light-field display

Since the microlenses are vanishingly small compared with the main lens, the main lens is effectively fixed at the microlenses’ optical infinity.

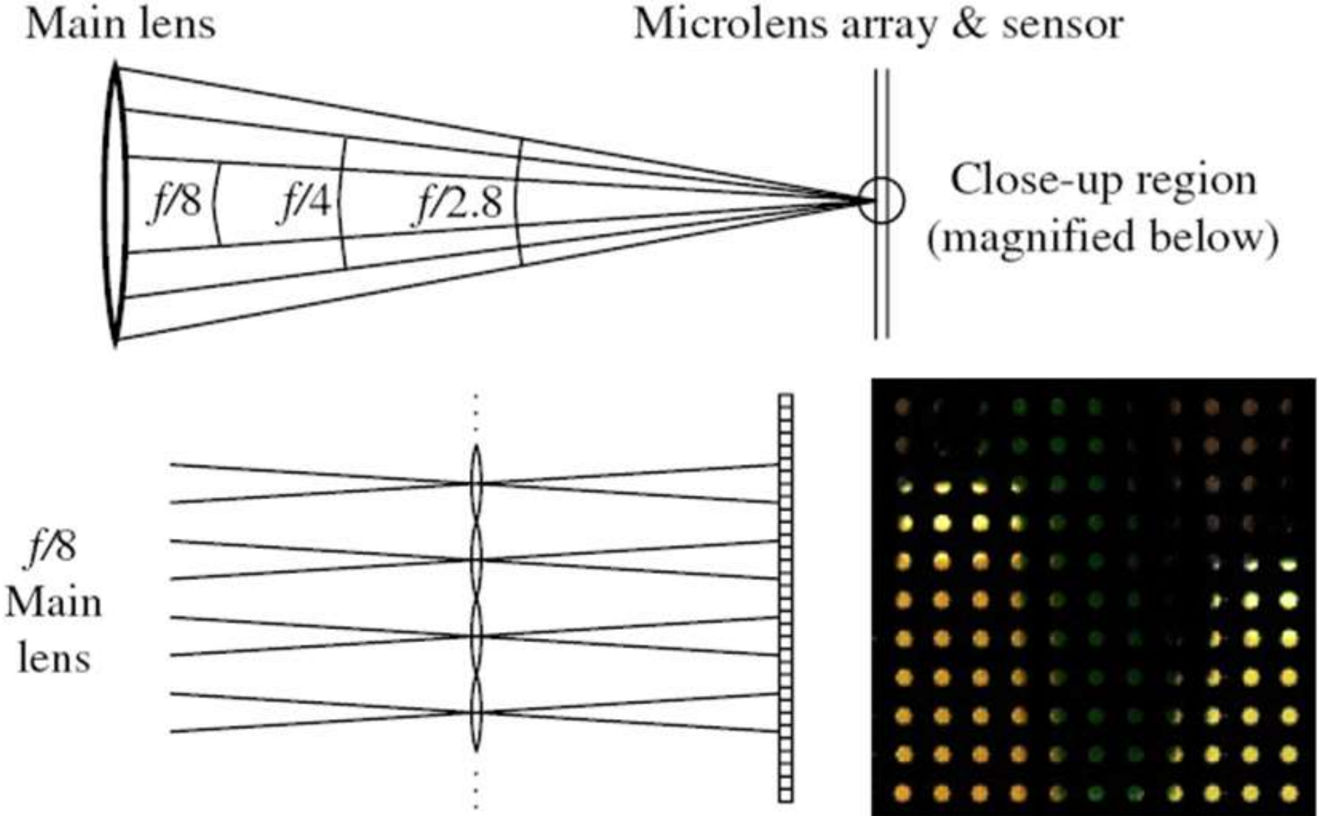



主鏡頭
u
微透鏡陣列
x
感光元件

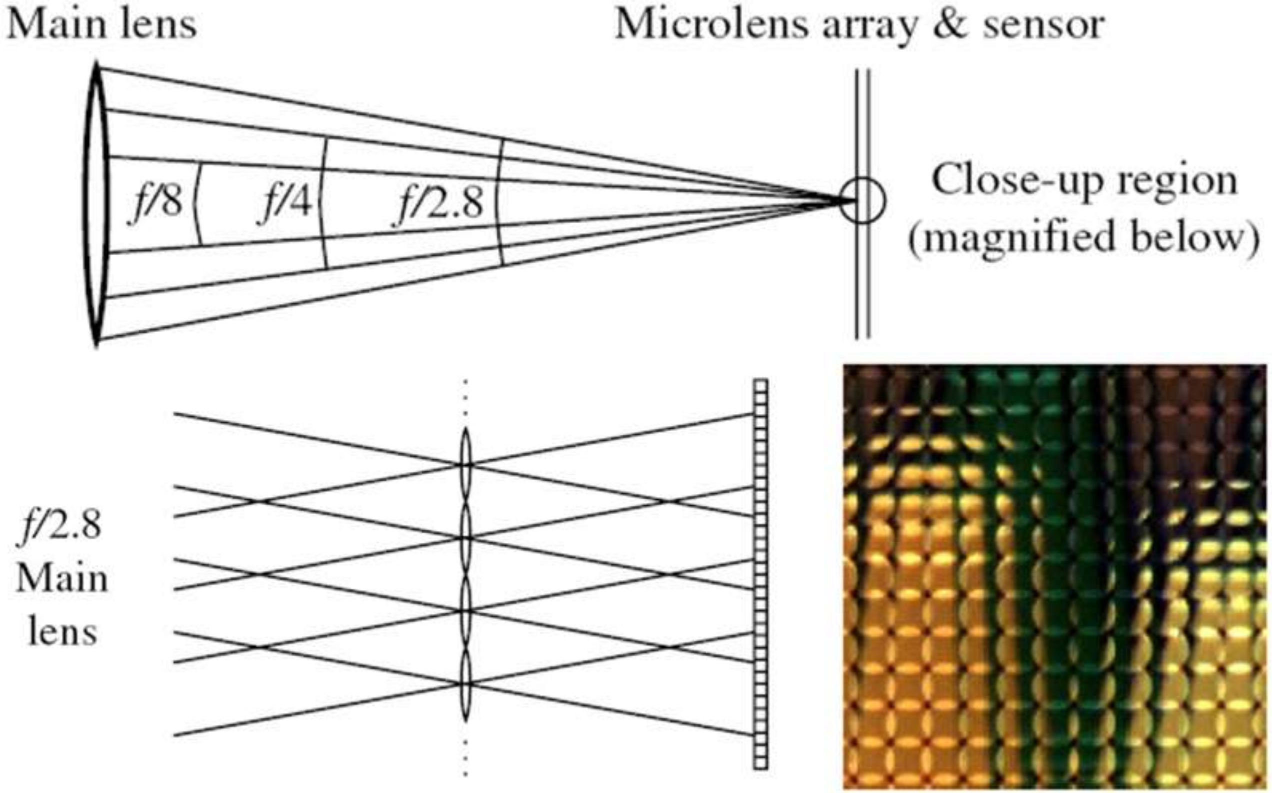

主鏡頭
u
微透鏡陣列
x
感光元件


主鏡頭
u
微透鏡陣列
x
感光元件

Matching Main Lens and Microlens f-Numbers



Matching Main Lens and Microlens f-Numbers



Matching Main Lens and Microlens f-Numbers

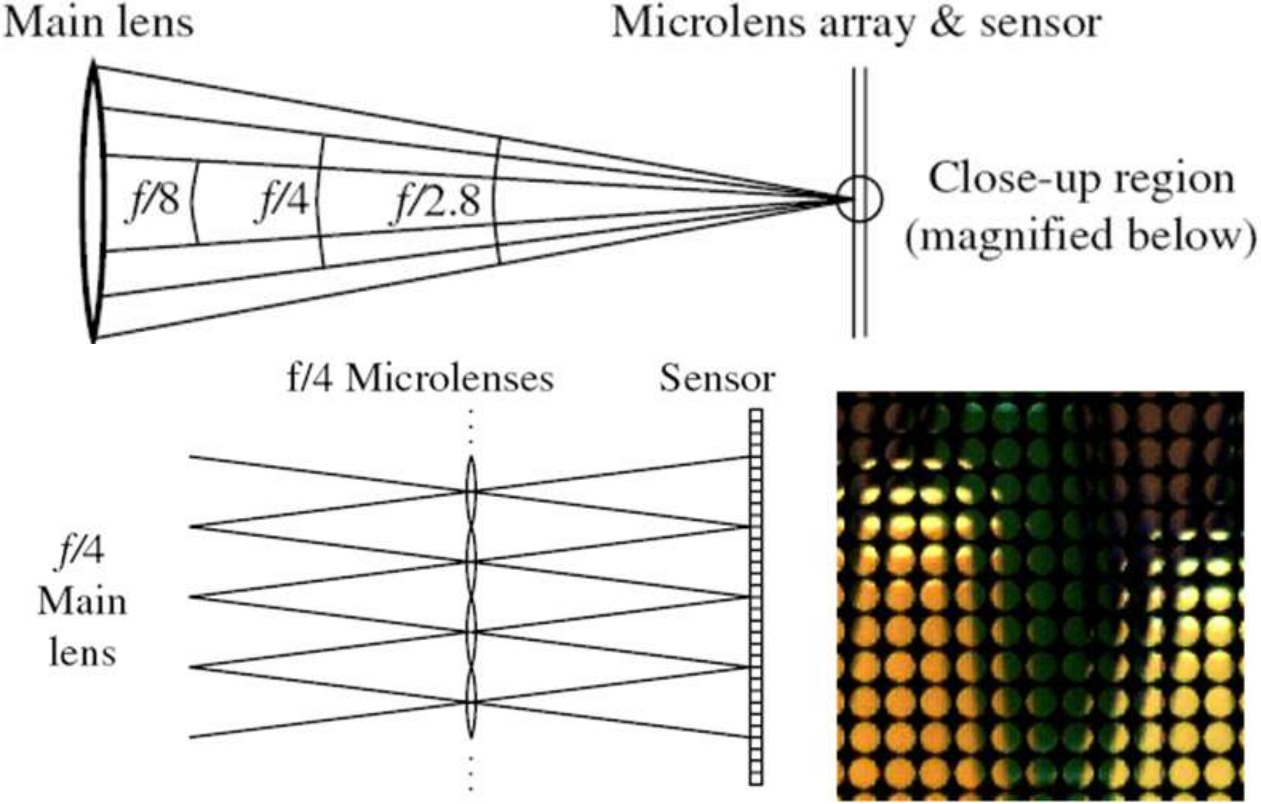


Image Synthesis

Two sub-aperture photographs obtained from a light field by extracting the shown pixel under each microlens (depicted on left). Note that the images are not the same, but exhibit vertical parallax.