Natural vision in AR

Whitepaper: CREAL's digital light field

Authors
Tomas Sluka, Alexander Kvasov, Tomas Kubes

Contact
info@creal.com

C•REAL

1. Summary

Augmented Reality (AR) is to become our everyday tool in everything from cooking to neurosurgery within this decade – the next big thing after smartphones, the natural next step in the development of communication technology. Yet, while most of the AR ecosystem evolves predictably, and gets incrementally better every year, the AR display technology needs a revolution.

Today's 3D displays provide conflicting depth information that causes adverse visual and neuro-ophthalmic effects - possibly including permanent damage to the eyesight - which may threaten the acceptance of the AR in the coming years.

This is why CREAL developed a display that cares for the user's vision, offering a digital image that supports the natural behaviour of the human eye. CREAL's light field display projects a highly efficient, high-fidelity digital representation of how light exists in the real world. This radically new type of display system provides correct focal depth to the digital imagery, seamlessly blending the digital and real worlds. This way, CREAL makes the 3D experience finally complete, natural and healthy.

This paper reveals the weakest link of today's AR displays, and explains the principles of CREAL's light field and its pros and cons in respect to other prospective AR display technologies.

AR flat display

- Visual conflict within arm's reach
- Eye-strain and nausea in <20 min
- Potential source of vision damage

The personal space is a no man's land for today's AR. Eyes cannot focus on real and virtual objects at the same time.

CREAL's AR light field display

- Life-like visual representation
- Extended use without conflicts
- Natural for human vision

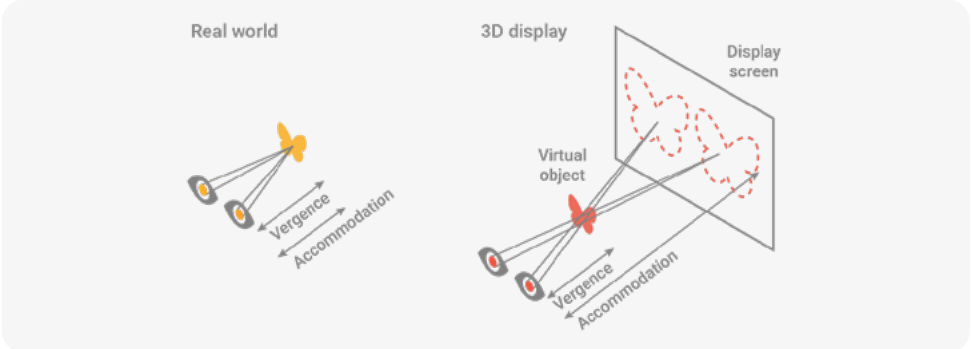
CREAL's light field displays virtual images in correct focal distances and brings AR within arm's reach.

2.
Flat display:
AR weakest link

Today's AR hardware creates 3D imagery without monocular (single eye) depth cues. Their absence causes at least three critically important visual conflicts that should be regarded and treated as a Trojan horse that we carry into the future of AR.

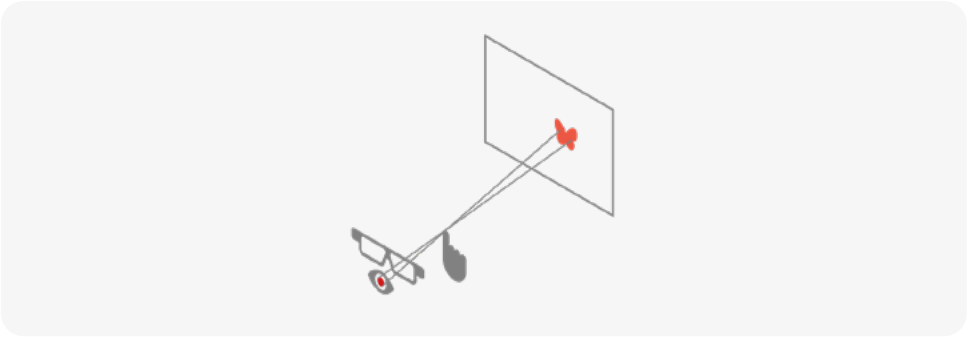
2.1 Vergence-Accommodation Conflict

Human eyes perform two motoric functions to perceive image depth: vergence (crossing of the two eyes) and accommodation (focus of each eye). These two eye functions work normally in sync, when they don't it is called the Vergence-Accommodation Conflict (VAC). Practically all today's VR/AR products use two flat image sources to imitate the stereoscopic illusion of image depth, but they entirely lack any focus depth and ocular parallax. The flat image sources support the vergence, but not accommodation forcing the viewer's eyes to focus at a wrong fixed distance. VAC causes eye-strain, nausea, and potentially even permanent damage to the eyesight [1–6].

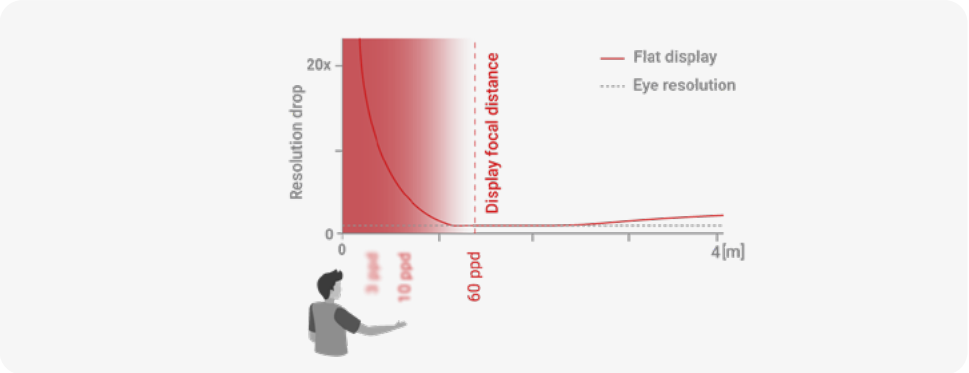


2.2 Focal rivalry

Flat fixed-focus images lead to an incorrect augmentation. Real objects have focal depth, virtual objects don't. For instance, it is generally impossible to display virtual objects in focus next to our own hands. This effect can be demonstrated with the display on which you are reading this text. If you close one eye and put a hand between the screen and you, you can try to "virtually" see **this text** sitting at the tips of your fingers. When the eye focuses on the fingers, however, the text becomes blurred and vice versa. It is impossible to see both in focus at the same time.

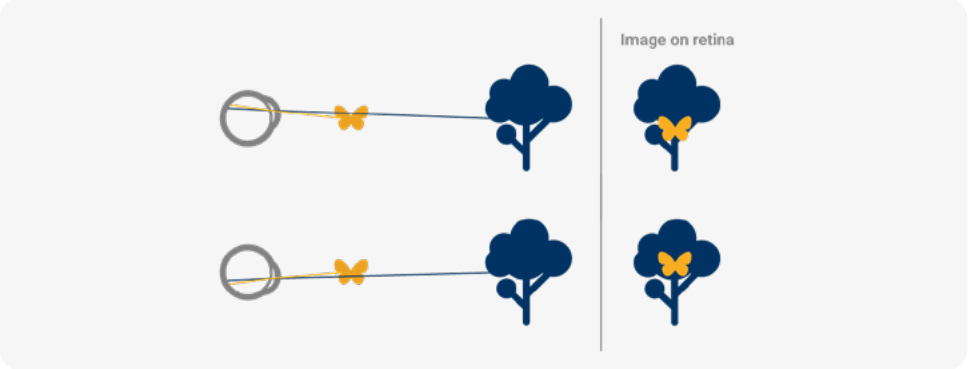


Typically, the focal distance of the flat displays in VR and AR today is set optically to 1.4 m, 2 m, or infinity. When the eye focuses at the distance of the display, the perceived resolution is limited only by the display resolution or by the classical limitations of the optics. The resolution of the eye at fovea is ~60 pixels per degree (ppd), or ~1 arcmin/pixel, which is satisfied by a full HD image displayed in ~20 deg Field of View (FoV). When the eye focuses at a different distance, however, the display appears blurred due to the eye defocus. This effect is extremely strong at distances below 1 m, approaching only 3 ppd at 20 cm. This makes the flat-display AR unusable in the personal space within arm's reach.

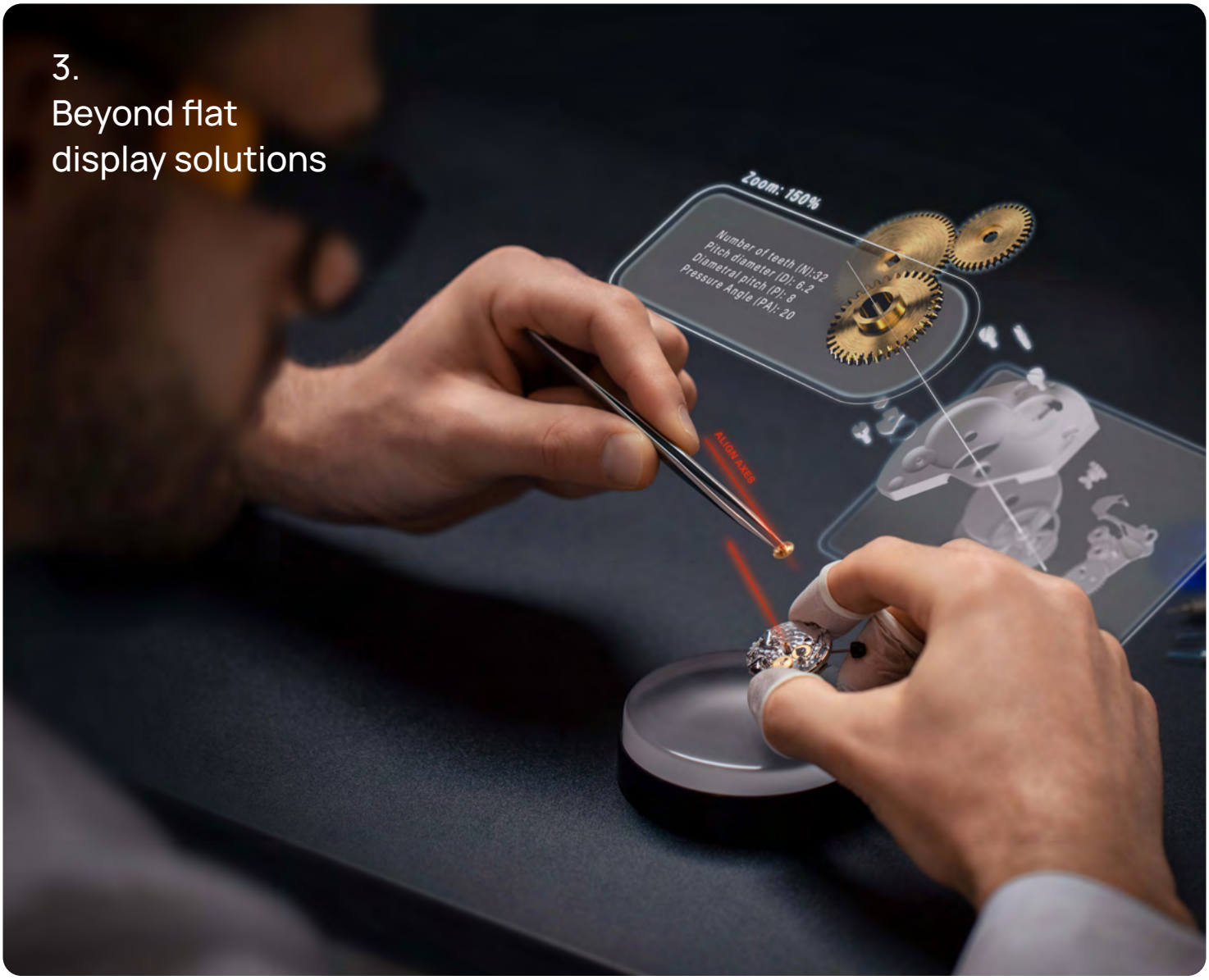


2.3 Ocular parallax

Small rotation of the eye in the eye-socket creates the so-called ocular parallax - perceived relative displacement between close and far objects. This depth cue, too, is absent in optically flat images, while it is arguably as important as the accommodation cues [7].



3. Beyond flat display solutions



3.1 Partial solutions

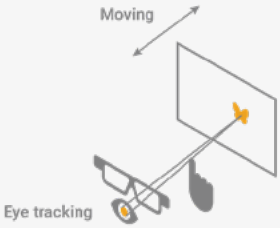
Tomorrow's AR depends on the paradigm shift in display technology today. No display on the market, however, provides fully satisfactory accommodation and ocular parallax cues. Only two partial solutions appeared in commercial devices.

3.1.1 Multifocal displays

Multifocal solutions place optically a flat display to multiple discrete focal planes. Either one depth plane at a time is used, based on the eye-tracking or content information, or the focal distance is rapidly cycled while each plane displays an image of only the optically closest virtual objects. The former approach was provided by Magic Leap with two depth planes, the latter by Avegant and LightSpace Technologies with four depth planes. The depth discretization is however limited and noticeable and the fast sweep is penalized by proportionally lower effective frame rate and brightness, and substantially higher complexity of the optics.

3.1.2 Varifocal displays

Varifocal solutions imitate the monocular depth cues with varifocal optical elements that move the focal distance of a flat display dynamically according to eye-tracking information. Their proper function requires digital imitation of the optical blur and ocular parallax. Eye-tracking has, however, inherently limited precision, response time, and reliability[8] while the imitation of especially the ocular parallax is possibly highly sensitive to it. The imitated depth cues may also provide unnatural visual input to observers with partly impaired vision who are used to certain, although imperfect, visual input. Overall, varifocal methods are critically dependent on eye-tracking, its calibration and proper imitation of monocular depth cues. It remains unanswered whether it can provide mid and long term advantages over more complete solutions discussed below.

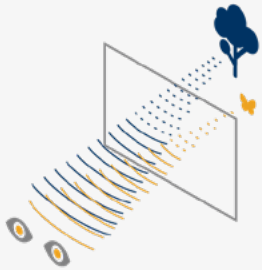


3.2 Complete solutions

Two concepts that provide technically more complete solution to all three problems above are in development:

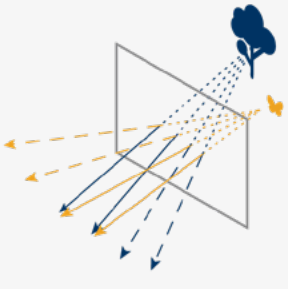
3.2.1 Computer-generated holography

Computer Generated Holography (CGH) uses Phase Spatial Light modulators to create discrete approximations of light waves. CGH is arguably the ultimate solution that can theoretically fully reconstruct light into the required form. In practice, however, CGH suffers from a range of optical and discretization artefacts, high sensitivity to temperature and driving voltage, low frame rate, and heavy computation requirements. Efficient high-fidelity CGH was not yet demonstrated even in laboratory conditions.



3.2.2 Digital light fields

Light field is a practical and simple approximation of real world light. Instead of treating light in terms of waves like CGH, it treats light in terms of rays or photons that bombard the eye from virtual points in space and build their image on the retina. Most of the realizations so far, however, provided low quality imagery. The rest of the document will be therefore about light fields, their current deficiencies, and how to make light fields work efficiently, well and now.



3.3 Light field solutions

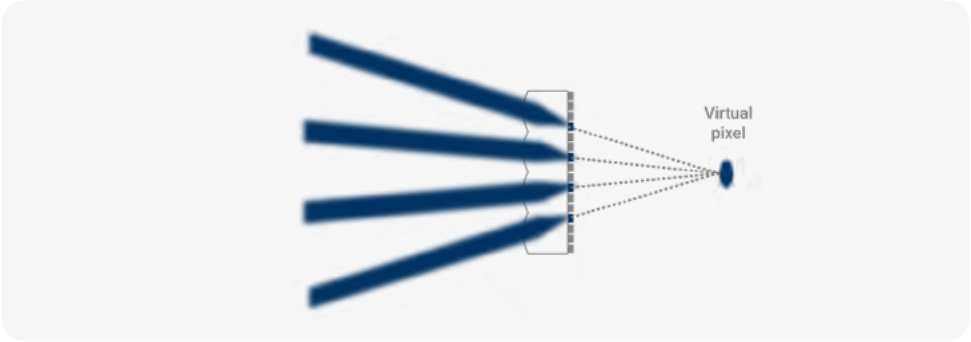
Real world light can be described as a continuous field of rays that are reflected, refracted, diffused or emitted by physical objects and propagate through free space. Each point in the real space transfers practically an infinite amount of light rays into an infinite range of directions. Digital light field is an engineering approximation of it.

3.3.1 “Classical” light field displays

A number of technical realizations of digital light field displays were conceived in the past, but the principle of all of them has a common base in “directional pixels”. Unlike classical display panels which emit the light/color of each pixel uniformly to “all” directions, directional pixels of light field displays project different colors (rays) into different directions. The array of beams emitted from an array of such pixels represents the digital light field.

Light field displays were traditionally constructed as modifications of classical flat displays with an attached lens array and corresponding transformation of the displayed image. Each lens collimates the light from individual pixels underneath into a fan of direction. The lenses, however, effectively split a higher resolution display into many lower resolution subdisplays.

Such spatially multiplexed light field displays are technically simple, but extremely inefficient and inaccurate. Each perceived virtual pixel is constructed by multiple real display pixels and the lens arrays provide fundamentally low quality collimation. Such light field displays require easily 30-80 times more data than a flat image to achieve comparable perceived quality - usually ending with substantially lower quality at comparable bandwidth. Several other systems, such as tensor displays, solved partial problems, but mostly inherited this fundamental inefficiency.



3.3.2 Practical light fields: All about efficiency

Brute force digital light fields as described above are rightly associated with an enormous amount of image data and still low quality imagery. In the following text, we explain how light fields can create high-fidelity imagery with comparable processing efficiency in respect to the classical flat imagery – creating focal depth at low computing cost.

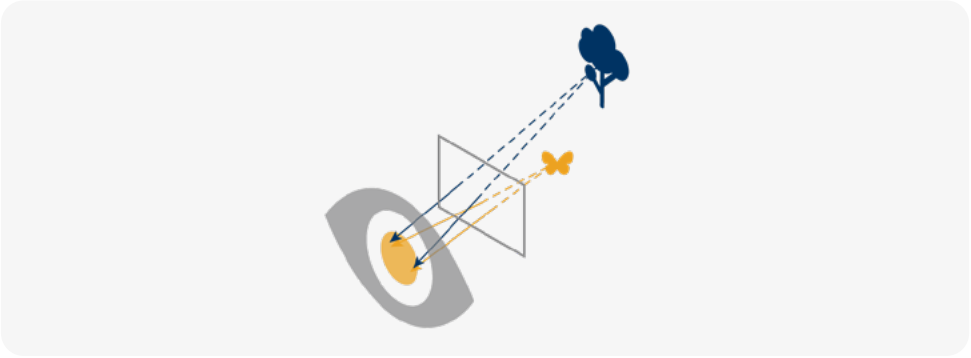
3.3.2.1 Near-eye projection

Classical TV-like light field panels project the vast majority of unique image data to no one’s eyes.

First step that drastically reduces the amount of light field data needed for the same perceived quality is to place the display near the eye where it can project all or majority of light field rays into the viewer’s pupil.

As a secondary consequence, FoV can be increased as it is not defined by a distant physical panel but by a near eye projection optics. Indeed, considering that the near eye display is moving with the head, FoV can be seemingly unbounded.

In contrast to flat displays and even to large light field panels, near-eye light field displays provide imagery with natural focal depth and, therefore, with correct monocular depth cues including the accommodation cues and ocular parallax.

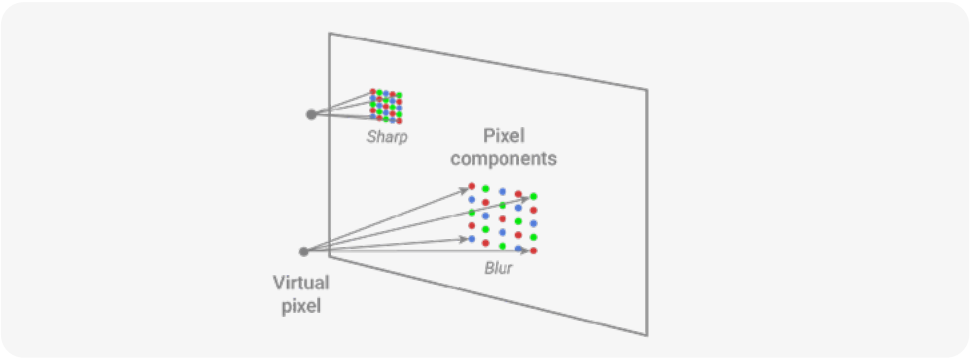


3.3.2.2 Data Efficiency

The brute force light field displays construct each virtual pixel from multiple real display pixels, often more than 20, leading to corresponding reduction of spatial resolution and redundancy in color resolution. That is an enormously bad trade-off.

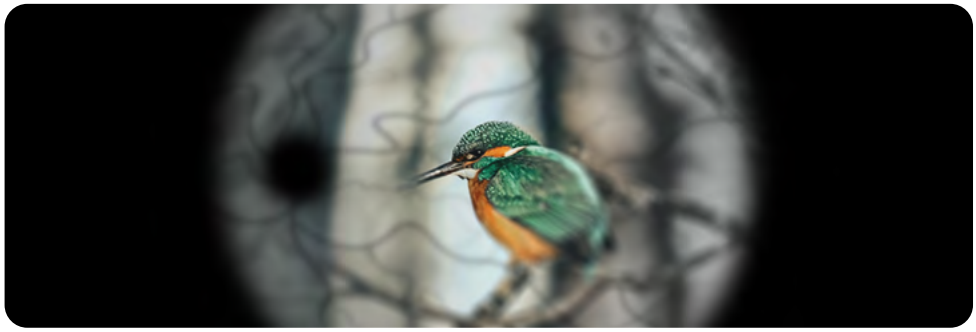
In a near eye projection system, however, individual rays from a virtual pixel do not need to carry unique and high color-resolution information because all (or most) rays recombine at the retina where they integrate the color of the virtual pixel, either in one point (in focus) or scattered (blurred).

It is therefore technically possible to render and transmit only the color and depth of a virtual pixel (depth is a byproduct of 3D rendering for 2D screens, too) and the color information can be distributed into an arbitrary number of rays at practically zero "creative" processing cost. Once this is performed on a hardware level, the bandwidth of a light field imagery is reduced to almost an equivalent of a flat imagery.

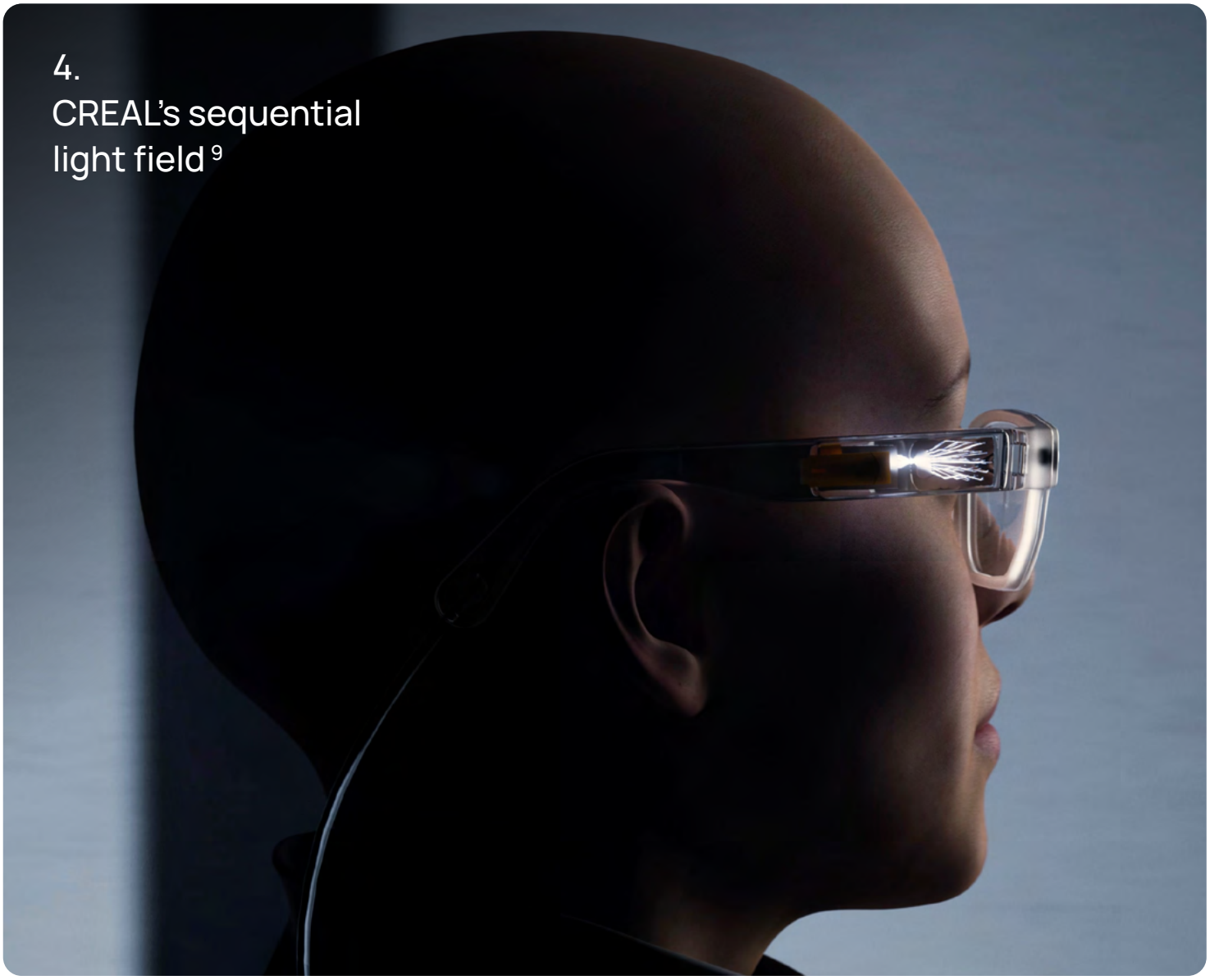


3.3.2.3 Foveation

An eye is a shockingly bad image sensor compared to what we seemingly perceive. Our brain is the amazing image processor that creates high quality visual sensation out of a very poor imaging input. An eye provides a high resolution image (~60 ppd) only at the so-called fovea that covers central ~5° FoV. That is like a coin size region half a meter away - the only part of the eye that can read this text. One is usually shocked when realizing that we can only read one or two words at a time without moving the gaze. The eye resolution rapidly decreases farther from the fovea up to the periphery where the eye is almost color blind and barely recognizes basic shapes of objects. When a display system matches the resolution distribution of the eye, the efficiency of image projection can be increased by orders of magnitude even compared to classical displays with uniformly distributed pixels.



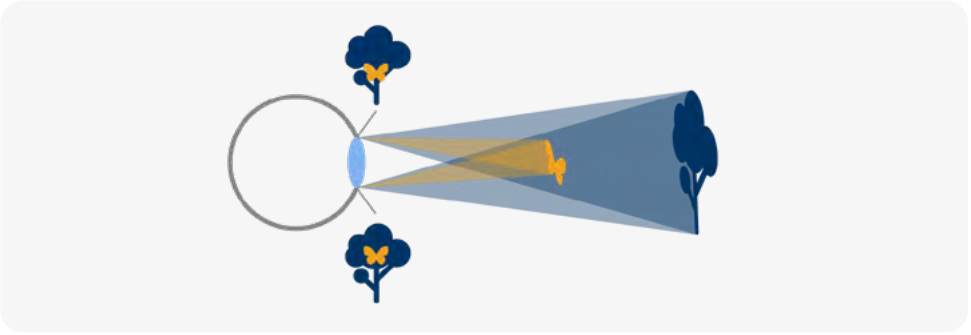
4.
CREAL's sequential light field [9]



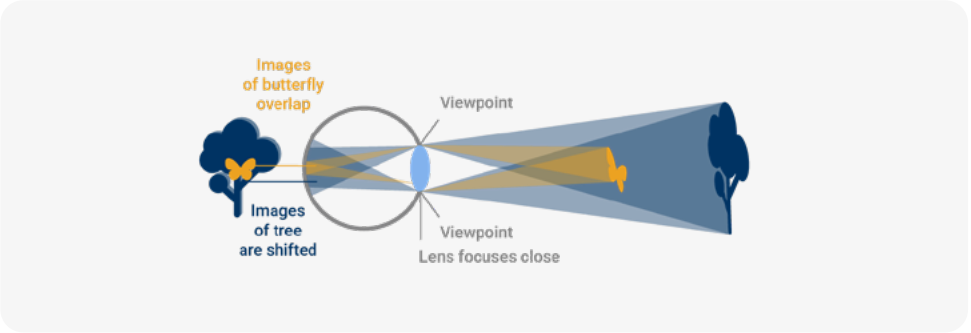
4.1 Principles

CREAL's near-eye display creates a highly efficient high-fidelity light field by projecting a fast sequence of images that represent slightly different perspectives of the same scene and that pass through an array of slightly displaced viewpoints to the eye pupil. Each of these images has high spatial resolution and low color resolution and is projected through a narrow aperture optics, similar to a pinhole, which causes that the individual images appear practically always in focus on the retina regardless of the actual focus of the eye. Together, however, they create a composed high-resolution image that is dependent on the eye focus as the individual always-in-focus images overlap and mutually move when the eye lens changes focus.

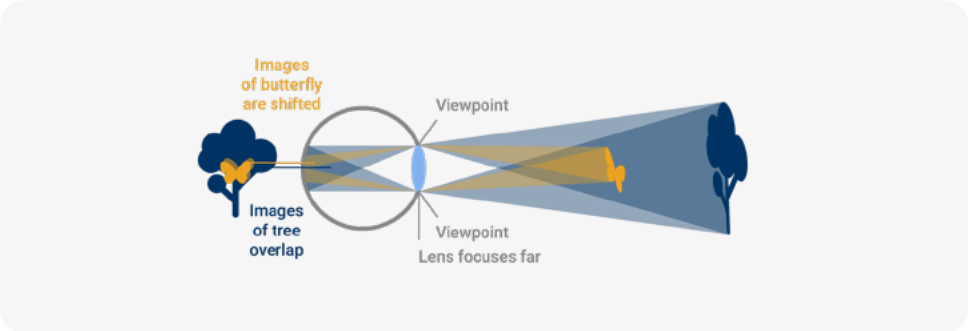
The below figures illustrate this mechanism with two viewpoints. If the light from a 3D scene enters the eye through two small apertures, two overlapping always-in-focus images appear on the retina. Both images display a butterfly and a tree, but the mutual position of the two objects in the individual images is shifted. For instance, in the image from the upper viewpoint the butterfly is lower in respect to the tree than in the image from the bottom viewpoint.



The eye lens then controls the mutual position of the images. If the eye focuses at the distance of the butterfly, the two images of the butterfly, each projected through a different viewpoint, overlap to create its sharp image, while the objects in another distances, such as the far tree, appear mutually displaced.

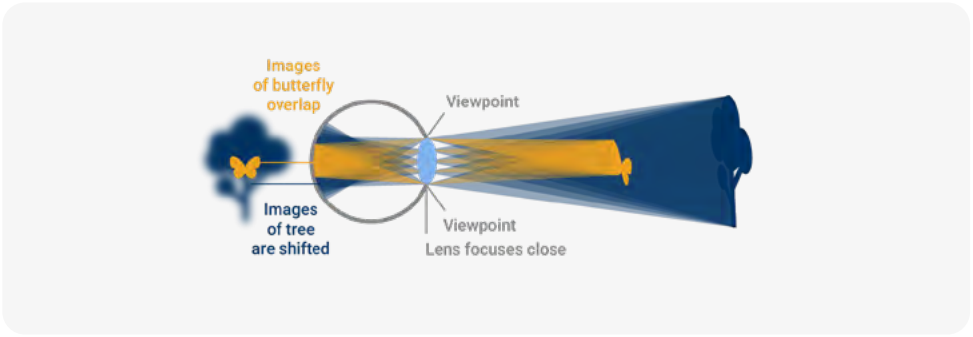


When the eye focuses at the distance of the tree, however, the two images on the retina shift and overlap to create a single sharp image of the tree while the butterfly doubles.



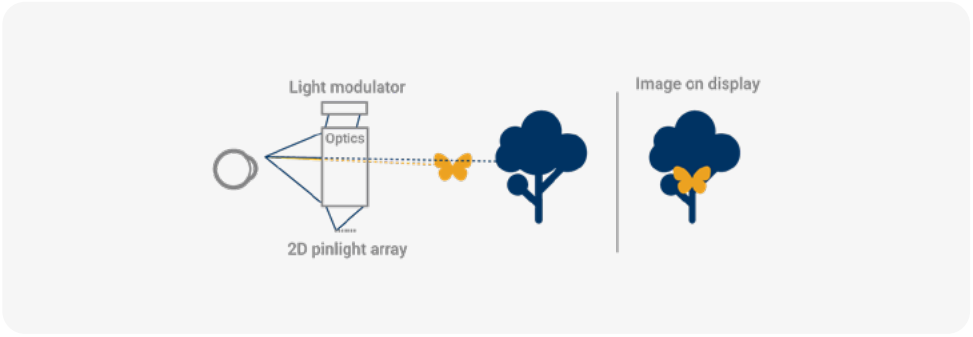
Since the two images on the retina can be continuously mutually shifted by the lens, the resulting image changes continuously as well. An object from any distance in the accommodation range can be therefore projected to overlap precisely on the retina and create a single sharp image. Thus, even two viewpoints project 3D imagery with optically infinite depth-resolution. In practice, the resolution is limited by the resolution of the individual images, optics or the eye. Nevertheless, with digital images, a unique depth plane can be defined as the configuration when the two images overlap on the retina pixel by pixel. Next depth plane requires that the lens shifts the images mutually by one pixel. For instance, if both images have 1000 pixels horizontally, the horizontal depth resolution will be 1000 depth planes spread from minus to plus infinity while the depth plane density will be highest at close proximity to the viewpoints/eye. Nevertheless, for most thinkable AR applications this represents an infinite resolution.

Obviously, the image doubling created by two viewpoints represents very unnatural blur. If the number of viewpoints is higher than two, (20 or more), the mutually shifted parts of the images on the retina appear as a smooth blur as illustrated in the figure below.

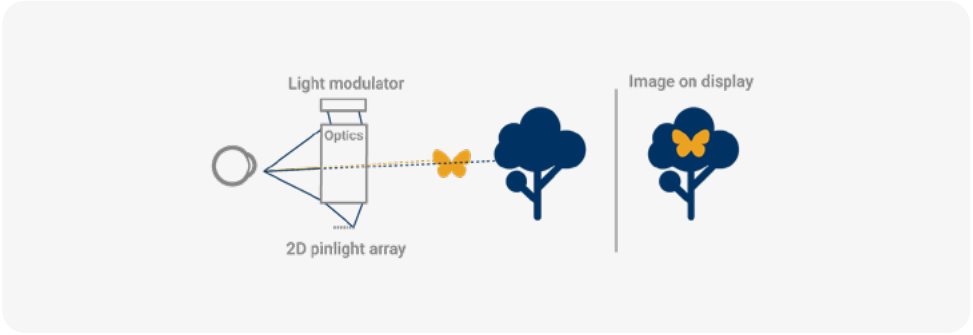


In the CREAL's system, the always-in-focus images are created by a sequential pin-light illumination of a fast spatial light modulator that reflects modulated light beams to imaging optics and towards specific viewpoints in the vicinity of the eye pupil. The modulator is technically a selective mirror that casts a shadow of the scene as it is supposed to be seen from the perspective of the particular viewpoint.

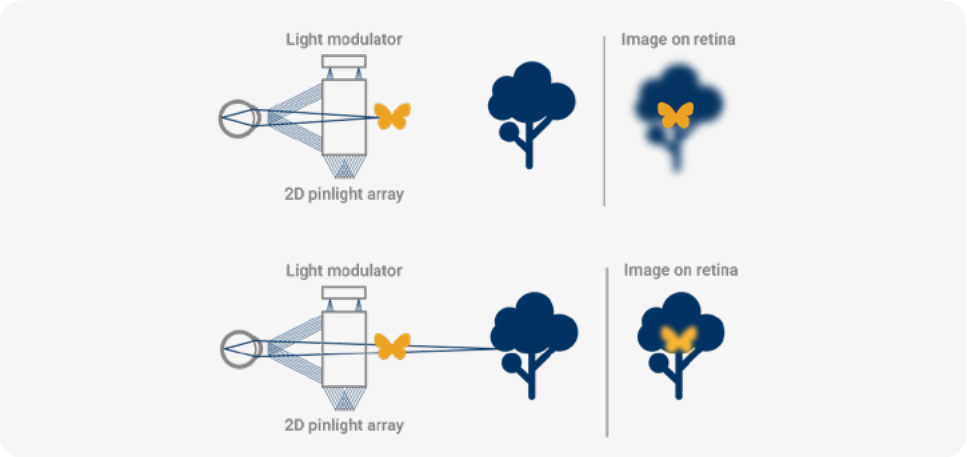
The small size of the pin-light-source (and Fourier filtering of diffracted light in the optical path) causes that the image has a large depth-of-field, i.e. the image is practically always-in-focus, and passes seemingly through a virtual pinhole hanging in the air near the eye pupil.



Different pin-lights perform the same operation in sequence, but the modulator reflects images of slightly different perspectives of the 3D scene and projects each through a different viewpoint.

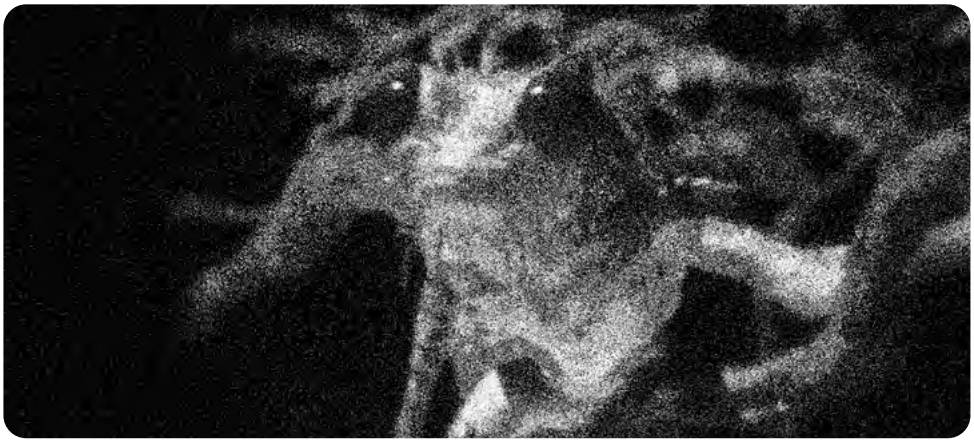


A fast sequence of always-in-focus images passing through a 2D array of viewpoints represents a light field that entirely or almost entirely enters the eye pupil. An eye can then focus on virtual objects in any distance. This operation is performed purely by the eye, no eye-tracking is needed. The light field was already reconstructed and has the properties of the real world light.



4.2 Light field components

Since the image that builds up on the retina is a sum of most or all of the projected light field components (90 to 180 always-in-focus images), each component can carry only fractional color information. The image on the right is a zoomed part of the most basic light field component. It is a binary image that, together with the other components, provides complementary information to the whole light field scene. The form of the individual light field components can be changed on the fly. For instance, the images can have higher color depth. The light field system can be made reverse compatible with classical stereo imagery as flat images are a subset of light field.



4.3 Performance

Following images are real photos of artificially generated light fields with the CREAL's system. Both photos capture the identical light field. Nothing was changed on the content or projection side between the shots. Only the camera changed focus. Both images consist of ~100 almost identical overlapping light field components. Each of the components passed through a different viewpoint near the camera entrance pupil. Only the camera focus determines when the images of the creature[10] overlap perfectly and construct a sharp image, while the images of the branches overlap only partly and appear blurred (left photo) and vice versa (right photo).

4.4 Optimal trade-off

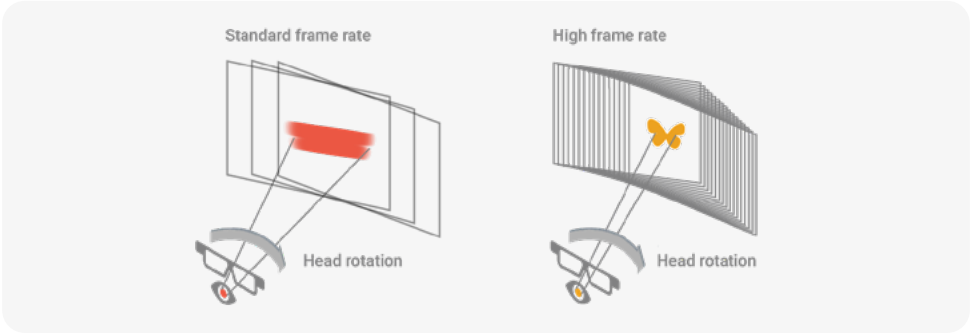
The sequential light field system forms the same basic elements as other light field systems (multiple images passing through multiple apertures), but, in contrast to the conventional systems, it achieves the optimum trade-off between spatial resolution, color resolution, number of viewpoints, FoV, eyebox size, and does not require any sensitive/complex optical elements such as micro lens arrays.

4.5 Digital lens and prescription corrections

When light field can be digitally formed, it can be also digitally transformed to apply arbitrary spherical, astigmatic or prismatic power, instantaneously and without any moving parts. Thus, light fields can digitally correct refractive errors of the imaging optics including the viewer's eye. No prescription inserts are needed, one click action can substitute it.

4.6 Natural light exposure

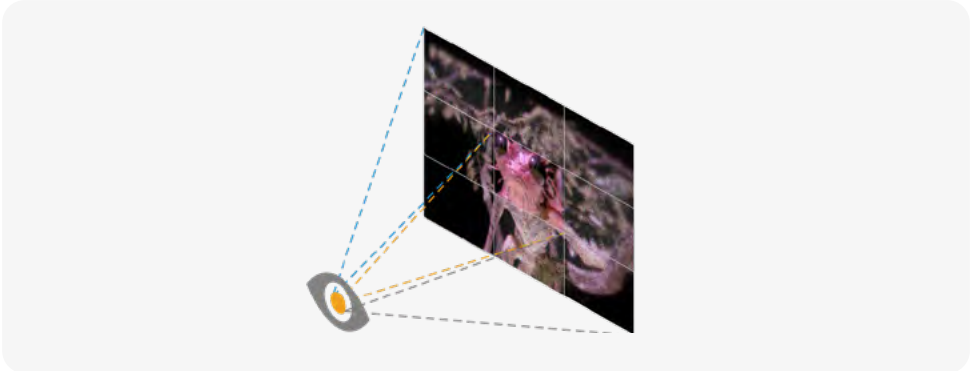
The fast sequence of low color-resolution images partly mimics the naturally continuous exposure of photons experienced by the eyes in the real world. CREAL's subframe rate (>6 kHz) is almost two orders of magnitude higher than that of classical display systems (30-120 Hz) eventually allowing to display fast moving virtual objects (if each component is rendered with a unique timestamp). The low frame rate of classical displays causes that each frame is displayed for a substantial amount of time when it seemingly moves relatively to the real world reference. If the base colors are sequenced at those low frequencies, colors appear mutually shifted and create a strong rainbow effect. High speed sequential light field minimizes this problem.



4.7 Highly efficient foveation

CREAL's light field projection system allows for individual optical and digital treatment of each light field component. Thus, it allows an ad-hoc distribution of the image information into different parts of the FoV (tiling), matching the non uniform resolution of the eye - both, spatial resolution and color resolution.

Such a foveated display can satisfy otherwise contradictory requirements of large FoV and large eye-box at the same time, high resolution light field at fovea and low color-resolution flat imagery at periphery, all with a low complexity projection system.



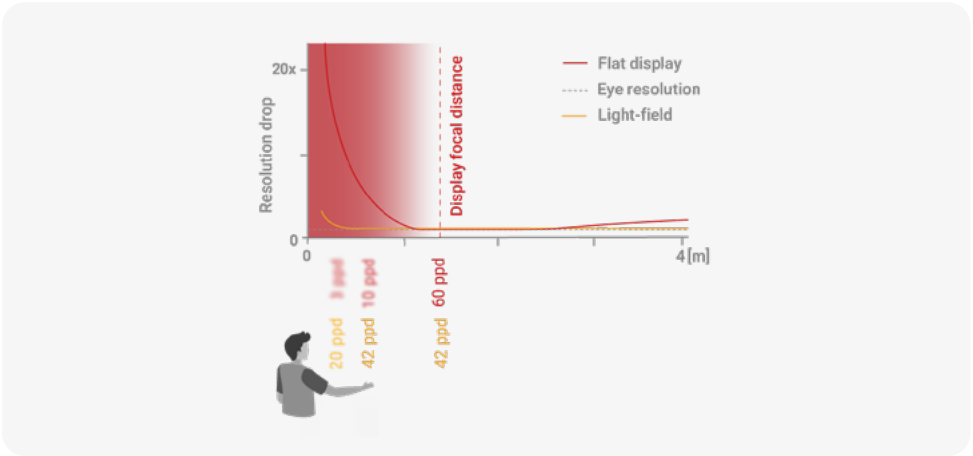
4.8 Summary of Benefits

- High spatial resolution (40-60 pixels/°)
- Practically unlimited depth resolution (>1000 depth planes)
- Correct monocular depth cues
- Refractive error corrections
- High data efficiency
- Very high subframe rate (>6 kHz)
- High contrast (>1000/1)
- Low complexity
- Mature base technologies

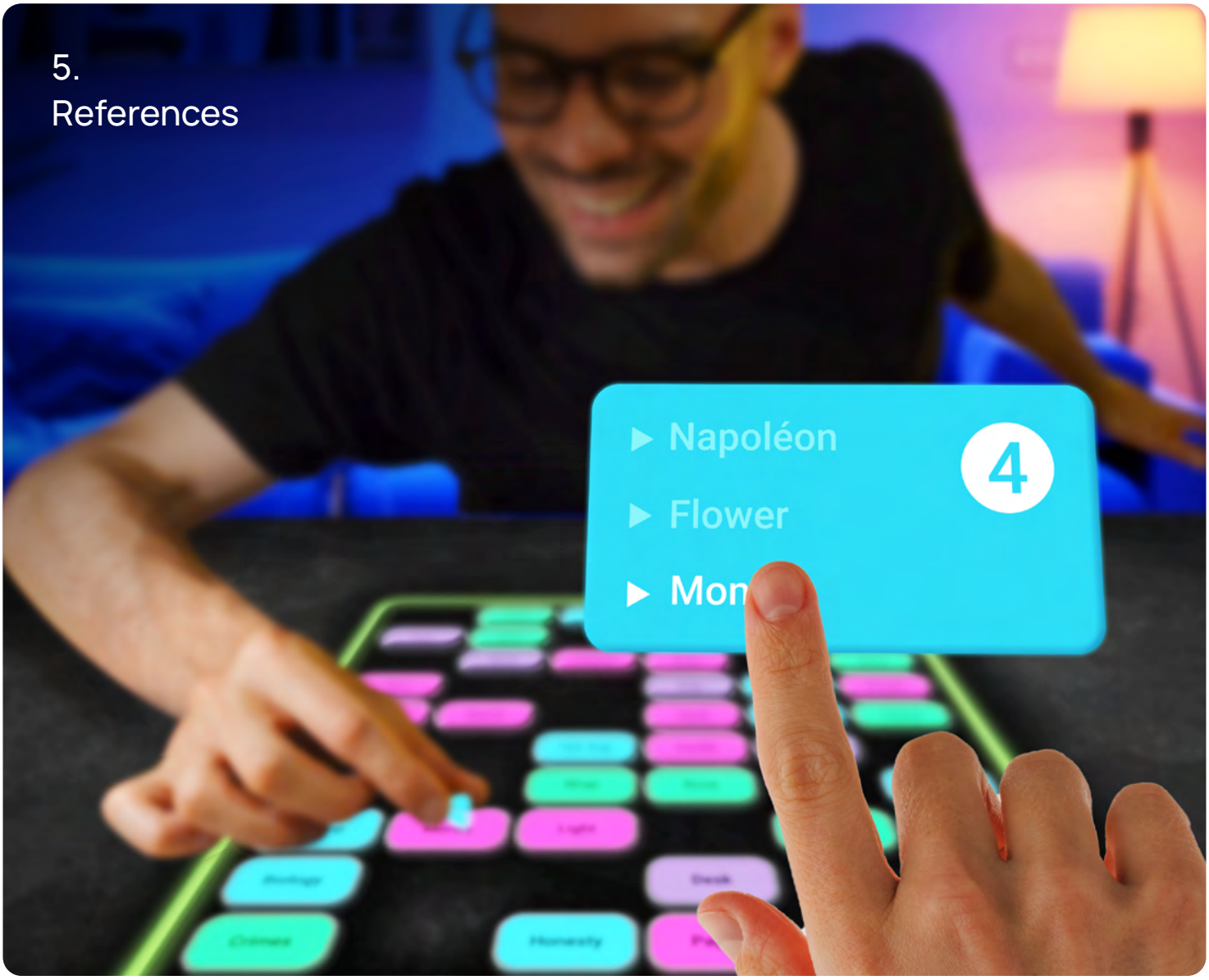
4.9 Drawbacks

Light fields have to balance a trade-off between two resolution limits. First is given by a self diffraction when the viewpoint aperture is too small. Second comes from the defocus due to the finite depth of field when the viewpoint aperture is too large. Like in the case of flat displays, only one focal distance can theoretically exceed retinal resolution (60 ppd) and the effective resolution drops farther from the optimal focal distance, but the drop is much smaller (to ~20 ppd at 5 diopters) compared to the resolution drop of the flat screen devices today (which goes to <3 ppd at 5 diopters). To put these limits into perspective: even the lowest light field resolution at the extremities of the accommodation range is comparable to the maximum resolution of today's AR/VR at their optimum focal distance.

Light efficiency of the amplitude modulator in the light field projector is lower compared to Laser Beam Steering (LBS), micro-LED, OLED, or CGH displays. The additional loss occurs at the modulator especially for sparse scenes because the whole modulator area is constantly illuminated while the "black pixels" damp the light. This loss must be compensated by a proportionally higher power budget for the light field illumination system. But is it really a considerable problem? How much power is actually needed? Even the brightest safe light which enters our eye pupils carries power in the range of microwatts. For an illustration, a smart-phone's or smart-watch's practically omnidirectional displays emit almost all their light to no-one's eyes, and yet they are still battery powered mobile devices. The critical bottleneck in the light efficiency of today's AR is not the display, but the optics. For example, the optical systems with diffractive combiners often have less than 1% efficiency and project the light to a large eyebox from which the eye pupil collects again less than 5% of light. Only <<1% of the emitted light then reaches the retina. This must be compensated by proportional boost of the light-source power which only then contributes considerably to the power budget and heating. Once the optical efficiency increases to even 5%, the power budget for light-source will be <10 mW at worst. This is an acceptable consumption for almost any realistic power budget which will be ultimately equally for all AR devices proportional to the information efficiency of image projection, image quality, and connectivity.



5. References



1. Hoffman, D. M., Girshick, A. R., Akeley, K. & Banks, M. S. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. Journal of Vision vol. 8 33 (2008).

2. Konrad, R., Padmanaban, N., Molner, K., Cooper, E. A. & Wetzstein, G. Accommodation-invariant computational near-eye displays. ACM Transactions on Graphics vol. 36 1–12 (2017).

3. Padmanaban, N., Konrad, R., Stramer, T., Cooper, E. A. & Wetzstein, G. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. Proc. Natl. Acad. Sci. U. S. A. 114, 2183–2188 (2017).

4. Lanman, D. EI 2020 Plenary: Quality Screen Time: Leveraging Computational Displays for Spatial Computing. https://www.youtube.com/watch?v=LQwMAI9bGNY (2020).

5. Microsoft documentation, Mixed Reality, Comfort. https://docs.microsoft.com/en-us/windows/mixed-reality/design/comfort (2020).

6. Developer warns VR headset damaged eyesight. BBC News https://www.bbc.com/news/technology-52992675 (2020).

7. Wetzstein, G. Computational Eyeglasses and Near-Eye Displays with Focus Cues (Conference Presentation). Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) (2020).

8. Banks, M. Are Leads and Lags of Accommodation Real? (Conference Presentation). Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) (2020).

9. Sluka, T. Near-Eye Sequential Light field Projector with Correct Monocular Depth Cues. WO2018091984

10. Credits for CGI design: Bystedt, D.: Tree creature. https://www.artstation.com/artwork/oL4Dq

CREAL.com |
info@creal.com |
EPFL Innovation Park, Switzerland