59th ILMENAU SCIENTIFIC COLLOQUIUM
Technische Universität Ilmenau, 11 – 15 September 2017
URN: urn:nbn:de:gbv:ilm1-2017iwk-141:2

SEE-THROUGH NEAR TO EYE DISPLAYS: CHALLENGES AND SOLUTION PATHS

Matthias Hillenbrand, Wolfgang Singer, Holger Münz, Norbert Kerwien

Carl Zeiss AG, 73447 Oberkochen, Germany

ABSTRACT

Many consider see-through near to eye displays the successors to the smartphone and envision a multitude of mixed reality and augmented reality applications. The ideal optical imaging system for a see-through near to eye display combines a large field of view (≥ 100°) with a large pupil (≈ 20 mm) and is both lightweight and unobtrusive.

In our contribution we first give an overview of challenges related to the design of see-through near to eye displays. Starting from the requirements of the human visual system, we then focus on two main performance parameters: field of view and aperture. These two parameters can be combined in a single parameter, the etendue. We show that the etendue of a see-through near to eye display is comparable to the etendue of lithography lenses and full frame camera lenses. To deliver the same etendue with a much lighter and more compact optical system is one of the main challenges of see-through near to eye displays. We discuss two possible solution paths: to increase the etendue close to the eye and to use foveated imaging concepts.

Index Terms - Augmented reality, near to eye displays, head-worn displays, pupil expansion, foveated imaging.

1. INTRODUCTION

Today's smartphones are masterpieces of engineering. They revolutionized the way we communicate and orient ourselves in the world. We even reached a point where smartphones can be used to augment the perception of the real world with virtual information. However, the field of view (FOV) covered by the smartphone is small, and camera-based see through has to be used to perceive the information behind the smartphone. Additionally, hands-free use cases are hard to realize. Imagine a time five to ten years into the future when we will be able to buy fully-immersive augmented reality glasses – i.e. glasses which combine unhindered see-through perception with a virtual image covering our full FOV. This vision led to massive investments in companies like Magic Leap [1]. Some experts even believe that the smart glasses may replace today's smartphones [2] [3]. Accordingly, the augmented reality (AR) market is expected to grow to $83 billion by 2021 [4].

The developers of AR devices still have to solve major technical hurdles before being able to build an unobtrusive AR device with large FOV and high brightness at a price attractive to consumers. In this paper we will explore selected challenges related to the design of the optical system and show possible solution paths. After an overview of technical challenges (Section 2), we will concentrate on the etendue requirements dictated by the human visual system (Section 3). The etendue is a parameter for the combined description of both the FOV and the system aperture. We will show that the etendue of an immersive AR system is comparable to the one of a lightography lens or a full frame camera lens. Such systems are far too bulky for head-worn applications. We thus show how the etendue may be increased close to the eye using multiple outcoupling (Section 4). Additionally we demonstrate, how foveated imaging concepts may be used to reduce the requirements imposed on an immersive AR system (Section 5). In Section 6 we will present our conclusions.

2. TECHNICAL CHALLENGES OF DESIGNING AN OPTICAL ENGINE FOR AR DEVICES

Building a successful AR system with large FOV and high resolution is an engineering challenge. First of all, we have to take the specific requirements of the human visual system into account. As humans, we constantly move our eyes to focus on different areas of our FOV. To make sure that the full virtual image is visible from all relevant eye pupil positions, we have to design systems with a large eyebox (eye motion box).

One major aspect of AR is to display virtual objects on top of a real world scene. The human visual system uses different depth cues to estimate the distance to an object. Two major cues are vergence (rotation angle of the eyes) and accommodation (object distance on which the eyes are focused). In many of today's AR and VR systems all 3D information is presented on a single virtual image plane, i.e. many of the 3D objects are presented at the wrong virtual image distance. This may lead to visual fatigue and possibly a failure of merging the stereoscopic images presented to the left and right eyes [5]. To avoid this vergence-accomodation conflict, the virtual 3D objects should thus be presented at different virtual image distances.

For maximum realism, they should also occlude the real world scene. Merely realizing occlusion is a technical challenge. The next technical challenge is related to the brightness and the resolution of the virtual image. To present a virtual image with a 100° diagonal FOV at a resolution matching the capabilities of the human eye (20/20 visual acuity), a display with more than 15 million pixels (16:9 aspect ratio) would be required.

Additionally, the brightness of the virtual image should match the one of the real world. However, the beam splitter which combines the real and the virtual image should only have a small impact on the see-through perception. This requirement typically results in a virtual image brightness which is significantly lower than the display brightness. Similarly, multiple outcoupling (see Section 4) leads to a reduction of virtual image brightness. In result, a display brightness of more than 30000 cd/m² may easily be required to realize a virtual image which is bright enough for outdoor use. Micro-displays fulfilling these resolution and brightness requirements are not yet available. This means that the AR development trends will be closely linked to the advances of micro-display technology.

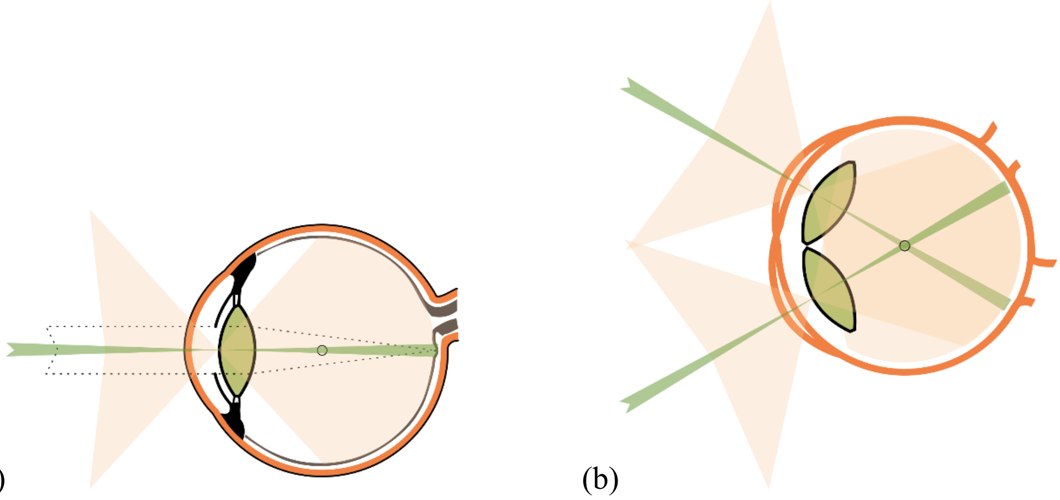
Apart from these challenges related to the optical system, there are many further challenges the developers of the full AR system have to face. These include: content, connectivity, user interface and user experience design, computing power, senor performance, battery life, as well as mechanical and thermal stability.

3. ETENDUE REQUIREMENTS OF THE HUMAN VISUAL SYSTEM

To present an image to the human eye, we have to adapt our optical systems to the specific requirements of human visual system. One of the most important properties is the varying density of photoreceptors across the retina. The highest density of cone cells (responsible for daylight and color vision) is found at the center of the fovea which spans a FOV of 5°. The density of cone cells reduces with growing distance to the center of the fovea. To perceive a large FOV at high resolution humans roll their eyes. This way, the total FOV is increased as well.

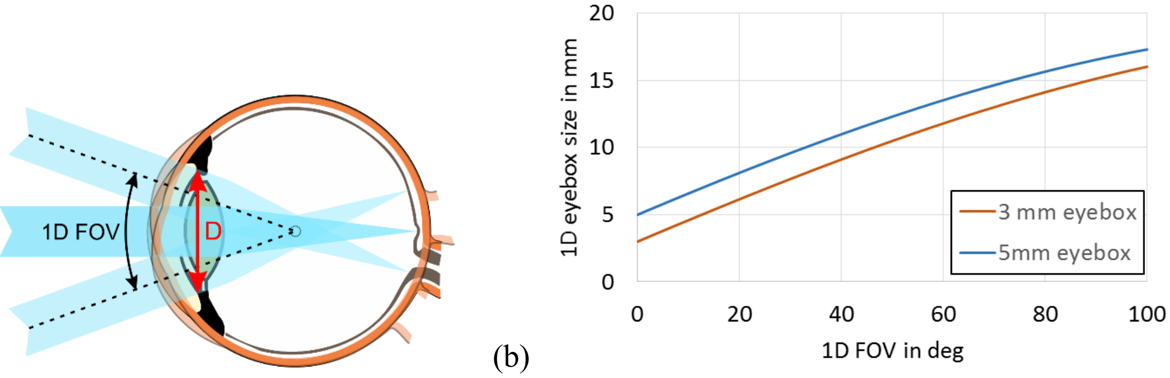
The FOV of the static eye is approximately 120° horizontally and 90° vertically. In contrast, the FOV of the rolling eye extends to approximately 165° horizontally (limited by the nose) and 150° vertically (limited by the eyebrows and the cheek bones). Fig. 1 illustrates the FOV of both a static and a rolling human eye.

To guarantee that all parts of the virtual image are visible from different rolling positions of the eye, the exit pupil of the optical system has to be significantly larger than the eye pupil. In visual systems a separate parameter – the eyebox – is used to describe the area within which the eye pupil may be moved while perceiving the full FOV, see Fig. 2(a). The dimension of the eyebox is typically larger than the theoretical movement range of the eye pupil to cover for alignment tolerances and different eye pupil distances.



(a) (b)

Figure 1: FOV of (a) the static and (b) the rolling human eye. Shown in blue is the area of sharp central vision while the pink and beige cones illustrate the peripheral FOV of the static eye. By rolling the eye, we can perceive the FOV between the two green cones shown in subfigure (b).



(a) (b)

Figure 1. (a) Definition of the eyebox length D by the movement range of the eye pupil and (b) dependence of D on the FOV.

According to Fig. 2(b) AR systems with large FOV also require a large eyebox. Both quantities, the FOV and the eyebox size (or the size of the exit pupil) can be combined in a single quantity – the etendue. The 1D etendue of an optical system is given by the following equation [6]:

$$G_{1D} = 2 \cdot NA \cdot L = 2 \cdot \sin\left(\frac{FOV}{2}\right) \cdot D \tag{1}$$

The 1D etendue is a quantity of conservation in ideal optical systems (without losses). Its paraxial approximation corresponds to the Lagrange invariant. In Table 1 the etendue of an immersive AR system with a FOV of 100° and an eyebox of 20 mm is compared to the one of classical imaging systems. This comparison shows that the etendue of an immersive AR system is comparable to the one of an immersion lithography lens or the etendue of a fullframe photo lens. Realizing an etendue this big with a low weight and unobtrusive optical system is one of the main challenges of see-through AR systems. In [7] it is shown that classical etendue-conserving imaging systems with the required etendue will be bulky and obtrusive.

To realize the required etendue in a compact and unobtrusive system we thus have to consider non-classical imaging systems. Multiple non-classical concepts can already be found in the literature. Without any claim of completeness, the solution space includes concepts like

- Multi-channel imaging [8].
- Multipe outcoupling [9].
- Use of diffusors or scattering plates [10].
- Dynamic exit pupil, e.g. by moving a scan mirror [11] or the optics [12].
- Transparent displays in combination with holographic collimators [13] or contact lenses [14].
- Foveated imaging [15].

Below, we will discuss two of these concepts in more detail: the etendue increase close to the eye by multiple outcoupling and foveated imaging.

Table 1. 1D etendue of different optical systems.

	Immersive AR	30x Ocular [16]	Immersion lithography lens /1900 [17]	Fullframe photo lens at F#2
Field size (long axis) in mm			26	36
FOV (long axis) in deg	100	68		
Numerical Aperture			1.35	0.25
Eyebox size (long axis) in mm	20	2.1		
1D Etendue in mm	31	2.5	70	18

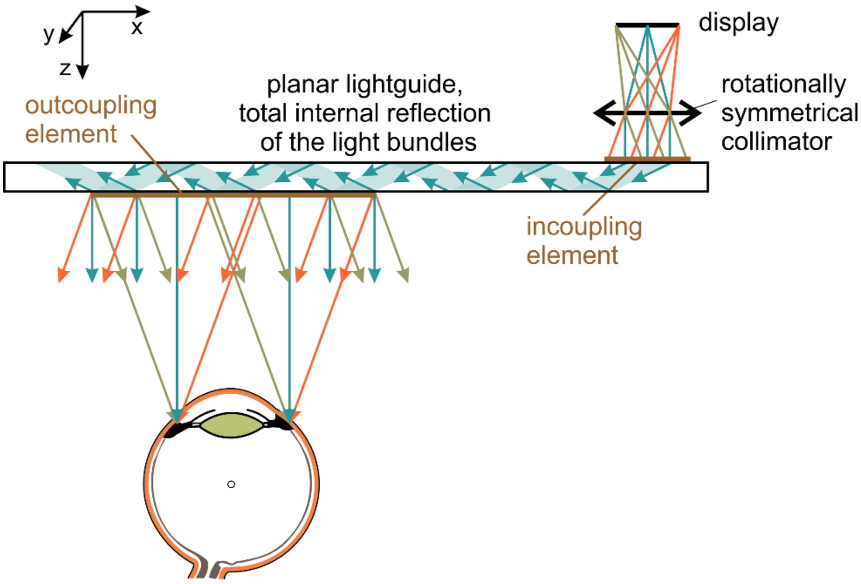


Figure 3. Etendue increase close to the eye by using a planar lightguide in combination with an outcoupling element which has low efficiency and is hit multiple times by the light propagating in the lightguide.

4. ETENDUE INCREASE CLOSE TO THE EYE BY MULTIPLE OUTCOUPLING

Fig. 3 illustrates the concept of multiple outcoupling. It is typically based on a planar lightguide which transports multiple collimated light bundles. Each light bundle is generated by a display pixel and is collimated by a rotationally symmetrical collimator. After collimation, the ray bundles hit the lightguide which includes an incoupling element. The incoupling element deflects the ray bundles so that they propagate within the lightguide through total internal reflection (TIR) until they reach the outcoupling element. At the outcoupling element the ray bundles are again deflected, leave the light guide and propagate towards the user's eye. The outcoupling element has a low optical efficiency, i.e. only a minor fraction of the beam is deflected while the major fraction of the beam is not deflected and propagates further within the lightguide. It thus hits the outcoupling element multiple times. Each time, a part of the ray bundle is coupled out. This way, a much larger bundle of parallel rays is generated and an extended eyebox can be achieved.

The major drawback of this approach is the reduced brightness (luminance) of the virtual image. The brightness reduction factor is given by the ratio between the exit pupil area of the collimating optics and the area of the outcoupling element hit by the corresponding ray bundle. The allowed expansion factor is thus given by the maximum display brightness divided by the required brightness of the virtual image.

For incoupling and outcoupling different technologies can be used. The possible options range from surface relief gratings and volume holograms to embedded Fresnel surfaces and slanted beam splitter arrays. Each technology has its own advantages and drawbacks e.g. with respect to mass fabrication and spectral selectivity. Fig. 4 demonstrates a major difference in deflection characteristics between gratings and slanted beam splitters. The deflection at a grating follows the grating equation

$$n \sin \varepsilon \pm n' \sin \varepsilon' = \frac{m\lambda}{p}. \tag{2}$$

The deflection at a mirror is governed by the law of reflection

$$\varepsilon' = -\varepsilon, \tag{3}$$

while the refraction at an interface between two media follows Snell's law

$$n' \sin\varepsilon' = n \sin\varepsilon. \tag{4}$$

Here, ε and ε' are the angles before and after deflection, respectively. Similarly, n and n' are the refractive indices before and after deflection, respectively. m is the diffraction order, λ is the wavelength, and p is the grating period.

For the case shown on Fig. 4 we assumed polycarbonate (n=1.59) as the lightguide material and considered a FOV of 20° outside of the lightguide. After refraction, the FOV is compressed to 12.6° within the lightguide. When deflected by the mirror (Fig. 4(b)), the angle between the deflected rays stays constant at 12.6°. The grating, however, introduces nonlinearities which result in an angle of 23° between the deflected rays. The angular range transported by TIR within the lightguide is the same for both cases. However, in the grating case of Fig. 4(a) this angular range is used less efficiently. In result, a smaller angular bandwidth can be transported in lightguides with grating couplers.

A second drawback of grating couplers is the dependence of the deflection angle on the wavelength. This means that different wavelengths of the same FOV are deflected into different directions. While the full FOV may be transferrable within the lightguide for the central wavelength, the FOV may be truncated for the outer wavelengths. To solve this issue, a stack of lightguides may be used. The Microsoft Hololens uses three lightguides, one for each band of the transmitted RGB-spectrum [18]. In contrast, a system like the Lumus DK-50, which is based on slanted mirrors, only require a single lightguide [19].

To transfer an extended FOV, the total FOV may be segmented into multiple channels. Each channel is transported in a different lightguide and the lightguides are again stacked on top of each other [20] [21]. The advantage of a larger field is thus bought at the expense of an increased total number of lightguides, more weight, lower yield, and a worse see-through perception.

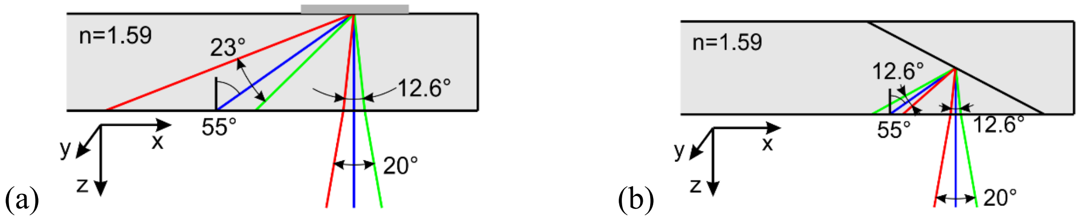


Figure 4. Comparison between the deflection at (a) a surface relief grating and (b) a slanted mirror or beam splitter. The different colors represent the chief rays for different field angles and are not related to the wavelength.

5. FOVEATED IMAGING

In Section 3 we pointed out that the visual acuity decreases with growing distance to the center of the fovea. Another important aspect of the human visual system is that we rarely use the full rolling range of our eyes. Instead, 85% of the saccades have rotation angles below 15° [22]. This enables foveated imaging systems with varying resolution across the FOV. Fig. 5 illustrates the resolution requirements across the FOV if a maximum eye rotation of 15° is defined. Outside of the high-resolution central region, the required resolution drops significantly.

Foveated imaging systems which take the properties of the human visual system into account can be built in various ways. The first option is to build a single channel imaging system with minimum complexity which offers high resolution in an extended central region and increasingly lower resolution towards the edges. A second option consists of multiple channels of the same type where the central channels are optimized for higher resolution [23]. Third, a manifold of options arises when different systems like those listed in Section 3 are combined. One possible option is the combination of a planar lightguide with slanted beam splitters in the center with a direct display concept in the outer region [15]. One challenge when combining multiple channels or multiple imaging concepts lies in combining the images in a way which does not distract the user.

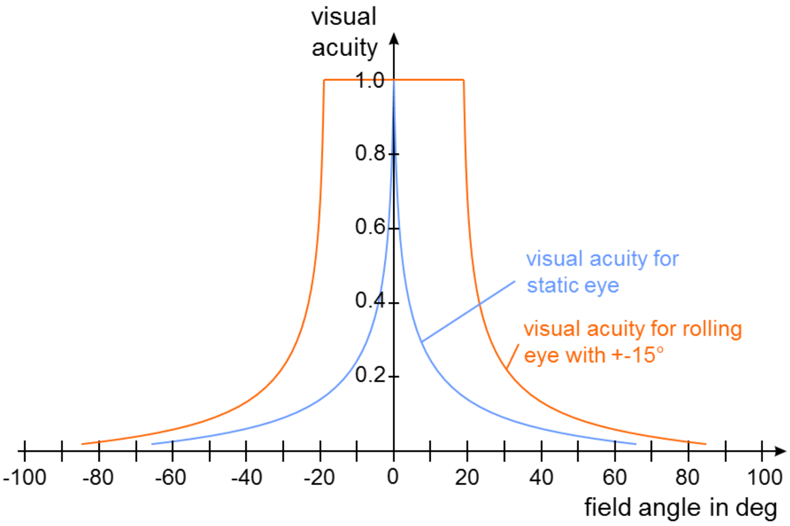


Figure 5. Visual acuity of the rolling eye with a maximum movement range of ±15°.

6. CONCLUSIONS

With a 1D etendue of approx. 30 mm, immersive but unobtrusive see-through near to eye displays are among the big challenges of optics. The vision of unobtrusive glasses which could at least partially replace the smartphone fuels massive investments in this technology.

In this paper we gave an overview of technical challenges linked to the design of an optical engine for an immersive see-through near to eye displays. We looked at one particular challenge, the presentation of a large FOV with a compact optical system. With multiple outcoupling and foveated imaging we discussed two possible solutions to this particular challenge and demonstrated that both solutions require specific tradeoffs. The finding that each solution brings its own tradeoffs can be transferred to many other aspects like the vergence-accomodation conflict and occlusion. Some solutions require emerging technologies like high-resolution micro-displays brightness and will thus take years to get ready for the market. At the same time there are many competing requirements and still mostly unsolved topics like occlusion. We thus believe that immersive and at the same time unobtrusive see-through near to eye displays will be masterpieces of engineering and will still take years to develop.

REFERENCES

[1] J. Gaynor, "magic leap announces $793.5 million in new funding," 02 02 2016. [Online]. Available: https://www.magicleap.com/#/press/magic-leap-series-c-funding. [Accessed 06 08 2017].

[2] M. Abrash, "How FB is Reinventing the Eyeglasses," 19 04 2017. [Online]. Available: https://www.youtube.com/watch?v=5Y-f45DsKmA. [Accessed 06 08 2017].

[3] J. Ward, "Microsoft HoloLens and Insider chiefs suggest smartglasses will replace smartphones," 04 08 2017. [Online]. Available: https://www.windowscentral.com/ar-smartglasses-may-replace-smartphones-hints-insider-chief-dona-sarka-and-hololens-creator-alex. [Accessed 06 08 2017].

[4] Digi-Capital, "After mixed year, mobile AR to drive $108 billion VR/AR market by 2021," 01 2017. [Online]. Available: http://www.digi-capital.com/news/2017/01/after-mixed-year-mobile-ar-to-drive-108-billion-vrar-market-by-2021. [Accessed 06 08 2017].

[5] D. M. Hoffman, A. R. Girshick, K. Akeley and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision,* vol. 8, no. 3, pp. 1-30, 2008.

[6] W. Singer, M. Totzeck and H. Gross, Handbook of Optical Systems, Volume 2: Physical Image Formation, Weinheim: Wiley, 2005.

[7] W. Singer, M. Hillenbrand and H. Münz, "Tackling the challenges of fully immersive head-mounted AR devices," *Proc. SPIE,* vol. ODC17, pp. 50-1 - 50-12, 2017.

[8] M. Higashihara, K. Inoguchi and M. Tohara, "Observation optical system". US Patent 9146399 BB, 25 09 2015.

[9] J. UPATNIEKS, "Compact head-up display". US Patent 4711512 A, 08 12 1987.

[10] M. Guillaumée, S. P. Vahdati, E. Tremblay and e. al., "Curved Holographic Combiner for Color Head Worn Display," *Journal of Display Technology,* vol. 10, no. 6, pp. 444 - 449, 2013.

[11] A. Stefan, M. Mailey, J. Choi and e. al., "Systems, devices, and methods for wearable heads-up displays". WO Patent 16134037 A1, 25 08 2016.

[12] S. Wakabayashi and M. Hino, "Image display device and drawing method". US Patent 2016124230 AA, 05 05 2016.

[13] lusovu, "Lisplay - how it works," [Online]. Available: http://lusovu.com/#/eyespeak. [Accessed 06 08 2017].

[14] R. Sprague, R. B. Sprague, H. Neidling, W. Meyers and J. Legerton, "Method and apparatus to process display and non-display information". US Patent 8441731 B2, 14 05 2013.

[15] J. S. Kollin and J. Lanier, "Hybrid display system". WO Patent 16204916 A1, 22 12 2016.

[16] ZEISS, "Victory DisScope Spektive," ZEISS, [Online]. Available: https://www.zeiss.de/sports-optics/de_de/natur/spektive/victory-diascope-spektive.html. [Accessed 06 08 2017].

[17] ASML, "TWINSCAN NXT:1980Di," [Online]. Available: https://www.asml.com/products/systems/twinscan-nxt/twinscan-nxt1980di/en/s46772?dfp_product_id=10567. [Accessed 06 08 2017].

[18] B. Kress, "Designing Mixed Reality Headsets," in *SID-ME 2017*, Dresden, 2017.

[19] Lumus, "DK-50," [Online]. Available: https://lumusvision.com/products/dev-kit-50/. [Accessed 06 08 2017].

[20] R. D. Brown, L. J. Burns, K. A. Stahl and R. B. Wood, "System for and method of extending vertical field of view in head up display utilizing a waveguide combiner". US Patent 8903207, 02 12 2014.

[21] P. Saarikko, "Exit pupil expanders with wide field-of-view". US Patent 8508848, 13 08 2013.

[22] T. A. Bahill, D. Adler and L. Stark, "Most naturally occurring human saccades have magnitudes of 15 degrees or less.," *Investigative Ophthalmology & Visual Science,* vol. 14, no. 6, pp. 468-469, 1975.

[23] P. Benitez, J. C. Minano and e. al., "Freeform optics for Virtual Reality applications," in *Design and Fabrication Congress 2017*, Denver, 2017.

CONTACTS

Matthias Hillenbrand	matthias.hillenbrand@zeiss.com
Wolfgang Singer	wolfgang.singer@zeiss.com
Holger Münz	holger.muenz@zeiss.com
Norbert Kerwien	norbert.kerwien@zeiss.com