Facial expression conveys nonverbal communication information to help humans better perceive physical or psychophysical situations. Accurate 3D imaging provides stable topographic changes for reading facial expression. In particular, light-field cameras (LFCs) have high potential for constructing depth maps, thanks to a simple configuration of microlens arrays and an objective lens. Herein, machine-learned NIR-based LFCs (NIR-LFCs) for facial expression reading by extracting Euclidean distances of 3D facial landmarks in pairwise fashion are reported. The NIR-LFC contains microlens arrays with asymmetric Fabry–Perot filter and NIR bandpass filter on CMOS image sensor, fully packaged with two vertical-cavity surface-emitting lasers. The NIR-LFC not only increases the image contrast by 2.1 times compared with conventional LFCs, but also reduces the reconstruction errors by up to 54%, regardless of ambient illumination conditions. A multilayer perceptron (MLP) classifies input vectors, consisting of 78 pairwise distances on the facial depth map of happiness, anger, sadness, and disgust, and also exhibits exceptional average accuracy of 0.85 ($p < 0.05$). This LFC provides a new platform for quantitatively labeling facial expression and emotion in point-of-care biomedical, social perception, or human–machine interaction applications.

1. Introduction

Facial expression serves as nonverbal communication that helps humans better perceive physical or psychophysical situations. Labeling facial expression or emotion is of much interest in clinical or healthcare applications, social perception, or human–machine interaction. Conventional facial expression reading uses physiological signals of electrocardiogram, electromyogram, or respiration, or facial features of 2D or 3D images (long-wave infrared, LWIR). Unlike 2D or thermal imaging, 3D imaging delivers stable facial topographic change despite face orientation, poses, and ambient light conditions. In particular, light-field cameras (LFCs) exhibit not only compact configuration due to the simple integration of microlens arrays (MLAs) between an objective and an image sensor, but also precise depth reconstruction due to small occlusion areas. As a result, the LFC attracts much attention for making it suitable for facial expression reading.

Conventional LFCs suffer from optical crosstalk between neighboring microlenses, resulting in degrading image contrast. Crosstalk-free MLAs utilize black silicon structures or optical absorbers and still have technical limitations in blocking broadband incident light. A high-contrast LFC using metal–insulator–metal optical absorber has been recently introduced and still has some technical limitations in optical design for 3D facial imaging. In addition, undesirable shadowing by indoor illumination conditions causes nonuniform and low-contrast facial images, leading to severe reconstruction errors. Multispectral or near-infrared (NIR) wavelength cameras either recognize human face with the spectral information or detect 3D mask face in uncontrolled illumination conditions, but they use a bulky system configuration with additional external sensors or lighting units.

Recent light-field imaging techniques are mainly focused on face recognition for presentation attack detection also called anti-spoofing, using feature extraction or machine learning. They either simply extract facial feature descriptors from raw light-field datasets or utilize visual disparity of different
perspective views or in- and out-of-focus of digital refocusing of light-field image at a local region of interest in the face such as ears, iris, or lower jaw. More recently, they also reconstruct face models based on model-free reconstruction of trained 3D face networks. However, they still struggle in quantitatively extracting 3D features for precise facial topography from the entire face.

Here we report facial expression reading from a precise depth map using NIR-based high-contrast and illumination invariant LFC (NIR-LFC) and machine learning classification. High-contrast and illumination invariant light-field images of human face are captured by NIR-LFC with optical crosstalk-free MLA and NIR bandpass filter (Figure 1a). The NIR-LFC exhibits a broad light absorption range and adopts a Keplerian configuration intentionally to achieve successful facial imaging. Under NIR illumination of the entire face, the light field passes through an objective lens, NIR bandpass filter, and MLA. The reverse and repeated microimages are finally captured on a single CMOS image sensor. The raw light-field image reconstructs high-accuracy facial depth map. In addition, the pairwise Euclidean distances between key facial landmarks from the depth map serve as input vectors \( (d_{p_i-p_j}) \) to a multilayer perceptron (MLP) classifier (Figure 1b). The output layer finally returns four different classes of facial expressions such as happiness, anger, sadness, or disgust.

2. Results

2.1. NIR-LFC Design and Microfabrication

The NIR-LFC is fully packaged with compact modules of NIR illumination and LFC (Figure 2a). The camera module comprises a single objective lens, inverted MLA with asymmetric Fabry–Perot filter (iMLA-AFF), narrowband NIR filter, and a single CMOS image sensor. The illumination module also includes vertical-cavity surface-emitting lasers (VCSEL, V102C121A-850, Osram, 850 nm in wavelength, 2 mW) and 3D-printed housing. The NIR bandpass filter (SLB850, SHHO, 800–900 nm) allows illumination invariant 3D facial images from various ambient illumination conditions. The farthest object distance of NIR-LFC is determined by considering the interpupillary distance of human face and the size of reconstructed light-field data, corresponding to 240 mm. The iMLA-AFF includes MLAs and AFF of chromium (Cr)–silicon dioxide (SiO2)–Cr. The AFF highly absorbs broadband illumination light from visible (vis) to NIR region and thus completely suppresses the optical crosstalk between adjacent microlenses. The 3D-printed housing of heat-resisting ABS resin and a printed circuit board for VCSEL sources are mounted to the LFC module (Figure 2b). The wafer-level microfabrication of iMLA-AFF is done using Cr lift-off, plasma-enhanced chemical vapor deposition (PECVD) of SiO2, photolithography, and thermal reflow (Figure 2c, extended...
microfabrication steps, and descriptions in Figure S1, Supporting Information). The scanning electron microscope (SEM) of iMLA-AFF shows hexagonally packed MLAs of 30 μm in diameter and 0.82 in fill factor (Figure 2d). The optical image of a fully packaged NIR-LFC shows the physical dimension of 8.4 mm × 14 mm × 5.6 mm (Figure 2e).

2.2. High-Contrast Light-Field Imaging

The iMLA-AFF plays a crucial role for high-contrast light-field imaging. Thin Cr layer shows optically lossy and less dispersive characteristics, suitable for both the top thin-film layer and the bottom reflector. The intermediate nondispersive SiO₂ layer exhibits high transparency in the vis—NIR region. Thus, the iMLA-AFF exhibits broadband absorption thanks to the extremely low Q-factor of cavity construction based on the strong interference effect.[30,31] The calculated thicknesses of the 6 nm-thick Cr film, 90 nm-thick SiO₂, and 140 nm-thick Cr film show the maximum absorption in the whole vis—NIR region, based on the finite difference time domain (FDTD) method (See Figure S2, Supporting Information). The optical sectioning and the line intensity profiles are compared for iMLA with and without AFF (25 μm in microlens diameter and 30 μm in pitch), using a confocal laser scanning microscope with an external laser diode (LM850103D6-CU, SHHO, λ₀ = 850 nm, 5 mW) (Figure 3a).[32] The normalized line intensities clearly show that the iMLA-AFF (line A–A’) completely eliminates the optical crosstalk compared with the iMLA (line B–B’). The raw images of line pairs captured from an USAF1951 target also substantially contribute to the image contrast after light-field reconstruction (Figure 3b). The Michelson contrast, that is, C = (I max – I min)/(I max + I min), of reconstructed light-field data from the NIR-LFC with AFF (blue line) is 0.66, which shows 2.1 times higher image contrast than that without AFF (red line). As the microlens diameter of NIR-LFC decreases, the number of pixels inside the microimage linearly decreases (angular resolution decreases), while the magnification of LFC, that is, the reciprocal of the size of an object projected onto a single pixel, further exponentially decreases (Figure 3c). The MTF50 value for subaperture images increases as the microlens diameter decreases, but it decreases at below 30 μm in microlens diameter due to the low magnification (Figure 3d, more information on MTF curves in Figure S3, Supporting Information) (Quick MTF software). The black short dashed line indicates a nonlinear Lorentzian fit for measured MTF50 values (black squares). As a result, the microlens diameter of NIR-LFC is set to 30 μm, which shows the maximum MTF50 value of 16.2 cycle mm⁻¹ and the angular resolution of 27 × 27 pixels. The raw images and their corresponding reconstructed light-fields of human

Figure 2. Configuration and microfabrication step of NIR-LFC. a) The NIR-LFC includes a compact objective lens, NIR bandpass filter, iMLA-AFF, and CMOS image sensor. The VCSEL sources (850 nm in wavelength) are assembled to NIR-LFC. b) Captured images of fully packaged NIR-LFC with VCSEL illumination. VCSEL housing is made from 3D printing of heat-resisting ABS resin (semitransparent) considering the heating of VCSELS. c) Microfabrication step of iMLA-AFF and camera module packaging. d) A SEM of iMLA-AFF. e) A photograph of a fully packaged NIR-LFC. The total physical dimension of camera module is 8.4 mm × 14 mm × 5.6 mm.
2.3. Illumination Invariant Facial Depth Map Estimation

The NIR-LFC captures the raw image of human face on a CMOS image sensor (Sony IMX 219, 2464 × 3280 pixels), which reconstructs the depth map after the light-field image calibration (Figure 4a). The raw image of \( I(s, t) \) is initially calibrated to assign the center position of individual microimages and then converted to the 4D radiance arrays of \( I(u, v, s, t) \) with the light-field imaging toolkit (LFIT v2.4).\[^{13}\] The disparity map is estimated using a cost volume-based stereomatching algorithm,\[^{16}\] which measures the similarity between the center view and the sub-aperture images and matches the cost of different disparity labels to estimate the stereo-correspondences with sub-label precision. In addition, the camera matrices, that is, intrinsic \([K]\) and extrinsic matrices \([R|t]\), are calculated from a raw checker board image with a known pattern size using the geometric calibration algorithm\[^{14}\] (more light-field reconstruction and depth map results in Figure S4, Supporting Information). Illumination invariant facial images are fully evaluated under different ambient illumination conditions using NIR-LFC and the 3D depth reconstruction process. White LED light (400–800 nm in wavelength) is illuminated on a human face at different illumination angles of 0°, 30°, and 60°, respectively (Figure 4b). Two VCSELs of a radiance angle of 60° / 45° also illuminate the entire face in front, 180 mm away from NIR-LFC. 13 different facial landmarks \( (P_0 - P_{12}) \) from reconstructed light-field data are extracted by using the OpenFace software.\[^{15}\]
Figure 4. Illumination invariant facial images through 3D depth maps acquired by NIR-LFC. a) 3D depth reconstruction processes from raw image. LFIT converts 2D raw image, \( I(s, t) \), to 4D light-field data, \( L(u, v, s, t) \). A disparity map obtained through cost volume-based stereomatching algorithm of the center and each subaperture image generates 3D depth map through disparity-depth conversion with camera matrices calculated using a geometric calibration. The depth map of male statue is 170 mm away from NIR-LFC is finally reconstructed through the disparity-depth conversion of disparity map and camera matrices. b) A schematic illustration of illumination invariant facial imaging using NIR-LFC. The NIR-LFC captures facial images for different illumination angles (0°, 30°, 60°) of white light LED (400–800 nm in wavelength). c) Light-field renderings and depth maps of facial images acquired at illumination angle of 60° with different NIR filters, that is, NIR cut filter (vis), no filter (vis+NIR), and NIR bandpass filter (NIR), respectively. d) Comparison of the depth map reconstruction error (RMSE) depending on the ambient light illumination angles under different NIR-LFC conditions. The NIR-LFC with NIR bandpass filter shows up to 54% lower reconstruction error than other NIR-LFCs, indicating that the NIR-LFC with NIR bandpass filter obtains a uniform depth map regardless of ambient light conditions.

and the nine different 3D Euclidean distances \( d_{ij} \) are then calculated from the depth map (more details of facial landmark extraction in Figure S5, Supporting Information). The facial landmarks of nose, eyes, and mouth are set to the distinct feature points. Note that the facial areas of cheek and forehead are easily saturated due to high specular reflection and thus excluded from the feature points. The illumination invariance of NIR-LFC with NIR bandpass filter (right [NIR], Figure 4c) at 60° illumination angle is compared with both NIR-LFC with NIR cut filter (left [vis]), Figure 4c) and NIR-LFC without any filter (center [vis+NIR], Figure 4c). The NIR-LFC with NIR bandpass filter captures uniform facial images despite the ambient illumination angles, and the reconstructed depth maps demonstrate clearly distinct features around the eyes, nose, and mouth. In contrast, other NIR-LFCs show severe shadows and shadings on the right side of the face, resulting in saturated and blurry areas in both the reconstructed light-fields and the depth maps. The reconstruction error from different Euclidean distances at each illumination angle is calculated by the root mean square error (RMSE) from the actual measured distance \( d_{ref} \) of ten subjects. (Figure 4d).

Note that the RMSE is defined as \( \sqrt{\sum_{i=1}^{n} (d_{ref,i} - d_{E,i})^2} \) \( (n = 9) \). The RMSE of NIR-LFC with NIR bandpass filter is nearly uniform up to 6.7 mm, whereas the RMSEs from other LFCs substantially increase with the illumination angle showing the RMSE of 12.8 mm or more at a illumination angle of 60°. The RMSE measurements clearly demonstrate that the depth map reconstructed by NIR-LFC with NIR bandpass filter significantly reduces the reconstruction error by up to 54%.

2.4. Facial Expression Reading of NIR-LFC

Machine learning-based facial expression reading is implemented using the facial features extracted from the facial depth map and the reconstructed light-fields. The NIR-LFC captures raw facial light-field images of 32 adult subjects. The subjects are asked to express an acted (or trained) facial expression, that is, happiness, anger, sadness, and disgust, to minimize external variables such as a gender, age, and ethnicity of subjects (Figure 5a). The total of 281 raw images consists of 38 images acquired by repeating the facial expressions once from 10 subjects and 243 images acquired by repeating the facial expressions three times from 22 subjects. Note that some images are excluded from the training dataset due to different external conditions such as illumination intensity, incident angle, and face orientation. The 281 datasets containing 78 Euclidean distance features from 32 subjects are augmented to 2,248 datasets by adding random noises from the original value of each feature to improve the robustness of the training data in the MLP classifier. The classification accuracy of each expression is calculated as the distribution of values obtained by 40 times 5-fold cross-validation.
validation from the entire datasets (Figure 5b, visualization of dataset distribution in Figure S6, Supporting Information). In all facial expressions, the classification accuracy of 3D facial features is 0.85 on average, substantially higher than that of 2D features. The p-values for 2D and 3D features according to facial expressions in the order of happiness, anger, sadness, and disgust are 0.043, 0.042, 0.01, and 0.043, respectively, indicating that the 2D and 3D accuracy distributions are significantly different from each other. c) Among the 78 features of each facial expression, top four dominant features in each expression are selected using the MI. 2D and 3D radar distributions of top four features for each expression clearly demonstrate that 3D facial features have distinct mutual dependency for each facial expression compared with 2D facial features. d) Dominant areas of face to read each facial expression based on MI values of all facial features. Highlighted areas in the face model contribute more than other areas to predict the corresponding expression.

Figure 5. Facial expression reading based on MLP classification from 3D depth maps and 2D images obtained by NIR-LFC. a) Reconstructed light fields and depth maps acquired from various facial expressions (happiness, anger, disgust, and sadness). b) In all expressions, the classification accuracy of 3D facial features is 0.85 on average, substantially higher than that of 2D features. The p-values for 2D and 3D features according to facial expressions in the order of happiness, anger, sadness, and disgust are 0.043, 0.042, 0.01, and 0.043, respectively, indicating that the 2D and 3D accuracy distributions are significantly different from each other. c) Among the 78 features of each facial expression, top four dominant features in each expression are selected using the MI. 2D and 3D radar distributions of top four features for each expression clearly demonstrate that 3D facial features have distinct mutual dependency for each facial expression compared with 2D facial features. d) Dominant areas of face to read each facial expression based on MI values of all facial features. Highlighted areas in the face model contribute more than other areas to predict the corresponding expression.
different weights to be attributed to the machine learning model for reading facial expressions. Therefore, the machine-learned NIR-LFC clearly demonstrates accurate facial expression reading using the MLP classification trained from precise 3D facial features.

3. Discussion

The NIR-LFC acquires high-contrast and illumination invariant face images and reconstructs them into accurate 3D depth maps. A broadband light absorption of AFF is mainly attributed to both the optical lossy nature and low dispersion of Cr. The interface reflection phase shifts become significantly different from 0 and \( \pi \) when a refractive index of the top thin Cr film has a comparable imaginary component \( (k) \) to real component \( (n) \) between 3.0 and 3.5 in the whole vis–NIR region.\(^\text{[36,37]}\) Therefore, the complex refractive index of Cr and the thickness of each Cr and oxide satisfying the critical coupling condition allow perfect absorption compared with other metals such as Au, Ag, and Al, which do not exhibit broadband absorption characteristics (Figure S2, Supporting Information). Meanwhile, the slight difference between the measured and calculated absorption of AFF using the FDTD method is related to a thin Cr film formation. The top Cr film thickness as low as 6 nm is randomly deposited as a form of both film and nanoislands during microfabrication to minimize the surface energy with the substrate, unlike assuming a uniform Cr film in the calculation. The random distribution causes less-than-0.1 absorption spectra difference in the corresponding region compared with the absorption spectra calculated by the FDTD method.

The depth reconstruction algorithm implemented takes 10 min for the 4D radiance arrays around 250 MB in size, which still limits a real-time 3D imaging. However, this can be significantly improved by parallelizing with GPU. Note that the computation equipment in this process uses an Intel i5 3.60 GHz CPU and 16 GB RAM. In addition, even after refinement process, including interpolation and filtering, specular reflection in no fine structure areas such as forehead and cheeks still interferes accurate depth reconstruction, which generally appears in stereo correspondence-based 3D reconstruction method.\(^\text{[16,38]}\) It is necessary to propose an alternative method of applying external illumination using a Moiré pattern or dot projection to LFC to overcome the issues. Despite a few limitations, the NIR-LFC obtains high-contrast and uniform 3D face images in various illumination conditions through an optimized optical design and optical crosstalk elimination from vis to NIR region.

The MLP classifier successfully reads facial expressions from 3D facial depth maps with high accuracy. The radar distribution and heat map from the MI values quantitatively evaluate the mutual dependency between the facial expression and the feature distribution. The results indicate that the specific features obtained from the 3D depth map play an important role in facial expression reading. Therefore, the 3D facial features from the NIR-LFC can further provide cognitive interaction information about weighted face positions in the processing of the human brain or machine learning when recognizing human expression.

4. Conclusion

In conclusion, this work has successfully demonstrated facial expression reading using compact, high-contrast, and a high-accuracy LFC and MLP classifier. The NIR-LFC comprises both NIR illumination module and camera module of a single objective lens, NIR bandpass filter, and iMLA-AFF on CMOS ISA. The total physical dimension is \( 8.4 \times 14.0 \times 5.6 \text{ mm}^3 \). The iMLA-AFF significantly increases the image contrast by 2.1 times and the NIR illumination module reduces the reconstruction errors by up to 54% regardless of ambient illumination condition. An MLP classifies the input vectors of 78 pairwise distances and clearly distinguishes four different facial expressions of happiness, anger, sadness, and disgust with exceptional average accuracy of 0.85 (\( p < 0.05 \)). This LFC provides a new platform for labeling facial expression reading and emotion in point-of-care biomedical, social perception, or human–machine interaction applications.

5. Experimental Section

Optical Design and Packaging of NIR-LFC: Each position of an image sensor, iMLA-AFF, and a single objective lens in NIR-LFC was numerically calculated considering the minimum total track length. The NIR-LFC intentionally selected the Keplerian configuration \( (f_{\text{fmla}} < B) \) due to the NIR bandpass filter thickness (0.4 mm in thickness) between the image sensor and the objective lens. The position of objective lens was 3.54 mm from the CMOS image sensor, considering the target depth-of-field range, MLA position (\( B \)), and the \( f_{\text{fmla}} \) (Figure S8, Supporting Information). Note that the objective lens’s focal length \( (f_{\text{obj}}) \), f\emph{MLA}, and \( B \) were 3.04 mm, 75, and 100 \( \mu \text{m} \), respectively. The theoretically calculated depth resolution of NIR-LFC was 1.7 mm (Figure S9, Supporting Information). An objective lens was mechanically separated from Raspberry Pi V2 camera first, and four spacer films were precisely attached to the edge of the image sensor and permanently bonded to the iMLA-AFF using a UV-curable adhesive. VCSEL housing and PCB were designed to be \( 8.4 \text{ mm} \times 14 \text{ mm} \times 2.9 \text{ mm} \) (\( W \times H \times D \)) and \( 8.4 \text{ mm} \times 4.7 \text{ mm} \) (\( W \times D \)), respectively, and completely combined with LFC module.

4D Light-Field Data Calibration and Depth Map Estimation: Raw microimages of NIR-LFC consisted of \( 2,464 \times 3,280 \text{ pixels} \). Using the calibration data file that displayed the center points of each microimage, indices \((u, v)\) represented the position of each microlens and indices \((s, t)\) represented the pixel within the microimage, respectively. Then, the raw images were reconstructed into 4D radiance arrays during the data calibration procedure according to the directional \((u, v)\) and spatial \((s, t)\) information using the LFIT.\(^\text{[31]}\) The 3D depth estimation algorithm was based on the phase shift in 2D Fourier domain for stereomatching between subaperture images with very narrow baseline.\(^\text{[16]}\) The disparity map was calculated by measuring the cost volume and by matching the similarity between the center and the subaperture images for estimating the stereo correspondences. Two complementary costs were used to match subaperture images, that is, the sum of absolute difference (SAD) and the sum of gradient differences (GRAD). The calculated disparity map was sequentially refined using multilabel optimization, iterative refinement process, and weighted median filtering of the cost slices. The final disparity map was estimated by user-defined parameters such as the number of pixels in a single microlens, the lens pitch, diameter, and a single pixel size. Simultaneously, a geometric calibration was conducted to subaperture images of a checker board with a known pattern size and extracted intrinsic \((K)\) and extrinsic \(([R|t])\) matrices. Each matrix contained focal length in \( x \) and \( y \) directions, principal point offset, and 3D rotation and translation information to elucidate camera geometric parameters. Note that all the processes were fully implemented in MATLAB.
MLP and MI for Facial Expression Classification: An MLP, one of the representative feed-forward artificial neural networks, learns the dataset by updating weights of connections of each neuron. A multilayer perceptron consisting of two hidden layers with 15 and 10 neurons was applied to classify various facial expressions with vectors acquired from 3D facial depth maps. Rectified linear unit (ReLU) was used as an activation function in the classification model. This model predicted one of four expressions (happiness, anger, disgust, and sadness) in the output layer. The model was further optimized by the ADAM optimizer with 3000 iterations during the training, and the nonconvergent models were excluded from the task. The dominant features to determine each facial expression were selected using MI.[15–19] MI between a feature and a facial expression measured how much the feature contributed to the expression reading. Highly scored features for each facial expression were selected from their corresponding MI.

Training Dataset Acquisition and Accuracy Calculation: The total of 281 raw images were acquired from 32 adult participants. For the experiment with human research participants, a written consent from all participants was obtained prior to the research including photos or reconstructed images of volunteers being taken and utilized as part of the research paper. KAIST Institutional Review Board (IRB) confirmed the written consent (approval number KH2020-105). 3D coordinates of 13 key facial landmarks (P0 – P12) were extracted from each image, and the pairwise distances (dij) between each landmark were calculated. 78 pairwise distances (dij) from one image were designated as a single vector data. Then, data augmentation was performed to enhance the robustness of the entire dataset by adding random feature noise dataset, which attributed variations to the original feature values. Note that the feature noise followed a standard normal distribution and the augmented data was generated by adding 1/10 of the average feature value of the original data and a random value from the corresponding distribution. As a result, 8 new images with added noise were generated per image, resulting in a total of 2248 augmented datasets.

The classification accuracy of each expression was calculated from the distribution of values obtained by repeating fivefold cross validation 40 times for the 2248 datasets. The randomly selected 4/5 dataset was used as a training set for the MLP, and the remaining 1/5 was used as a test set to calculate accuracy. The accuracy was a metric for how well the MLP classifier matched each facial expression, that is, number of correctly classified images for the expression data/number of images of the actual expression.

Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
This work was financially supported by the National Research Foundation, funded by the Ministry of Science and ICT (2021R1A2B5B03002428 and NRF2012M3A9C4048758) and Technology Innovation Program (20012464), funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea). The Interactive Supporting Information of this article can be found at: https://www.authorea.com/doi/full/10.22541/au.163482629.96595014.

Conflict of Interest
The authors declare no conflict of interest.

Data Availability Statement
The data that support the findings of this study are available in the supplementary material of this article.

Keywords
deep learning, facial expression reading, light-field cameras, machine learning, multilayer perceptrons, near-infrared imaging, 3D cameras

Received: September 20, 2021
Revised: October 29, 2021
Published online: December 16, 2021

[1] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, 2013 10th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG), IEEE, Piscataway, NJ pp. 1–6.
[2] A. J. Calder, A. W. Young, Nat. Rev. Neurosci. 2005, 6, 641.
[3] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J. G. Taylor, IEEE Signal Process. Mag. 2001, 18, 32.
[4] J. Selvaraj, M. Murugappan, K. Wan, S. Yaacob, Biomed. Eng. Online 2013, 12, 1.
[5] Y. Chen, Z. Yang, J. Wang, Neurocomputing 2015, 168, 871.
[6] D. S. Wickramasuriya, M. K. Tessmer, R. T. Faghih, in 2019 IEEE Healthcare Innovations and Point of Care Technologies, (Hi-POCT), IEEE, Piscataway, NJ pp. 9–12.
[7] K. I. Chang, K. W. Bowyer, P. J. Flynn, IEEE Trans. Pattern Anal. Mach. Intell. 2005, 27, 619.
[8] A. F. Abate, M. Nappi, D. Riccio, G. Sabatino, Pattern Recognit. Lett. 2007, 28, 1885.
[9] L. Yin, X. Wei, Y. Sun, J. Wang, M. J. Rosato, in 7th Int. Conf. on Automatic Face and Gesture Recognition (FG2006), IEEE, Piscataway, NJ pp. 211–216.
[10] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, R. Slama, IEEE Trans. Pattern Anal. Mach. Intell. 2013, 35, 2270.
[11] S. Jarlier, D. Delplanque, M. Delplanque, L. Cayeux, M. Velazco, D. Sander, P. Vuilleumier, K. R. Scherer, IEEE Trans. Affect. Comput. 2011, 2, 2.
[12] I. Kemelmacher-Shlizerman, R. Basri, IEEE Trans. Pattern Anal. Mach. Intell. 2010, 33, 394.
[13] G. R. Alexandre, J. M. Soares, C. A. Pereira, Pattern Recognit. 2020, 100, 107108.
[14] M. Martinez-Corral, B. Javid, Adv. Opt. Photonics 2018, 10, 512.
[15] T.-C. Wang, A. A. Efros, R. Ramamoorthi, IEEE Trans. Pattern Anal. Mach. Intell. 2016, 38, 2170.
[16] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, I. So Kweon, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, IEEE, Piscataway, NJ pp. 1547–1555.
[17] D. Keum, K.-W. Jang, D. S. Jeon, C. S. H. Hwang, E. K. Buschbeck, M. H. Kim, K.-H. Jeong, Light Sci. Appl. 2018, 7, 1.
[18] K. Kim, K.-W. Jang, J.-K. Ryu, K.-H. Jeong, Light Sci. Appl. 2020, 9, 1.
[19] S. I. Bae, K. Kim, K. W. Jang, H. K. Kim, K. H. Jeong, Adv. Opt. Mater. 2021, 9, 2001657.
[20] H. Chang, A. Koschan, M. Abidi, S. G. Kong, C.-H. Won, in 2008 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, IEEE, Piscataway, NJ pp. 1-6.
[21] J. Liu, A. Kumar, in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops, IEEE, Piscataway, NJ pp. 47–52.
[22] S. Z. Li, R. Chu, S. Liao, L. Zhang, IEEE Trans. Pattern Anal. Mach. Intell. 2007, 29, 627.
[23] R. Raghavendra, K. B. Raja, C. Busch, IEEE Trans. Image Process. 2015, 24, 1060.
[24] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, R. Ramamoorthi, European Conf. on Computer Vision, Springer, New York pp. 121–138.
[25] A. Sepas-Moghaddam, M. A. Haque, P. L. Correia, K. Nasrollahi, T. B. Moeslund, F. Pereira, IEEE Trans. Circuits Syst. Video Technol. 2019, 30, 4496.
[26] S. Kim, Y. Ban, S. Lee, Sensors 2014, 14, 22471.
[27] C. Galdi, V. Chiesa, C. Busch, P. Lobato Correia, J.-L. Dugelay, C. Guillemot, Sensors 2019, 19, 2687.
[28] A. Sepas-Moghaddam, F. Pereira, P. L. Correia, IEEE Trans. Inf. Forensics Secur. 2018, 13, 1696.
[29] M. Feng, S. Z. Gilani, Y. Wang, A. Mian, in Proc. of the European Conf. on Computer Vision (ECCV), Springer, 2018 pp. 501–518.
[30] Z. Li, E. Palacios, S. Butun, H. Kocer, K. Aydin, Sci. Rep. 2015, 5, 1.
[31] M. A. Kats, R. Blanchard, P. Genevet, F. Capasso, Nat. Mater. 2013, 12, 20.
[32] S.-I. Bae, K. Kim, S. Yang, K.-W. Jang, K.-H. Jeong, Opt. Express 2020, 28, 9082.
[33] J. Bolan, E. Hall, C. Clifford, B. Thurow, SoftwareX 2016, 5, 101.
[34] Y. Bok, H.-G. Jeon, I. S. Kweon, IEEE Trans. Pattern Anal. Mach. Intell. 2016, 39, 287.
[35] T. Baltrusaitis, P. Robinson, L.-P. Morency, in 2016 IEEE Winter Conf. on Applications of Computer Vision (WACV), IEEE, Piscataway, NJ pp. 1–10.
[36] M. A. Kats, D. Sharma, J. Lin, P. Genevet, R. Blanchard, Z. Yang, M. M. Qazilbash, D. N. Basov, S. Ramanathan, F. Capasso, Appl. Phys. Lett. 2012, 101, 221101.
[37] J.-B. You, W.-J. Lee, D. Won, K. Yu, Opt. Express 2014, 22, 8339.
[38] Z. Cai, X. Liu, X. Peng, Y. Yin, A. Li, J. Wu, B. Z. Gao, Opt. Express 2016, 24, 20324.
[39] B. C. Ross, PloS One 2014, 9, 87357.
[40] A. Kraskov, H. Stög Bauer, P. Grassberger, Phys. Rev. E 2004, 69, 066138.