IIEEJ Transactions on Image Electronics and Visual Computing

Special Issue on
Journal Track Papers in IIEEJ Annual Conference 2023

Vol. 11, No. 2 2023

The Institute of Image Electronics Engineers of Japan

Editorial Committee of IIEEJ

Contact Information

The Institute of Image Electronics Engineers of Japan (IIEEJ)
3-35-4-101, Arakawa, Arakawa-ku, Tokyo 116-0002, Japan
Tel : +81-3-5615-2893　Fax : +81-3-5615-2894
E-mail : hensyu@iieej.org
http://www.iieej.org/ (in Japanese)
http://www.iieej.org/en/ (in English)
http://www.facebook.com/IIEEJ (in Japanese)
http://www.facebook.com/IIEEJ.E (in English)

IIEEJ Transactions on
Image Electronics and Visual Computing
Vol.11 No.2 December 2023

CONTENTS

Published two times a year by the Institute of Image Electronics Engineers of Japan（IIEEJ）
3-35-4-101, Arakawa, Arakawa-ku, Tokyo 116-0002, Japan
Tel：+81-5615-2893 Fax：+81-5615-2894 E-mail：hensyu@iieej.org http://www.iieej.org/

Upon the Special Issue on Journal Track Papers in IIEEJ Annual Conference 2023

Editorial Committee

The 51st IIEEJ Annual Conference 2023 (Media Computing Conference 2023) was held in Toyama City between 28th and 30th August, 2023. The conference of this year adopted hybrid style (on-site and on-line) as same as the 50th meeting held on summer 2022 in Shiretoko, Hokkaido.

The conference site of Toyama was Toyama-Kenmin-Kaikan (Toyama Prefecture Citizen Hall), and two meeting rooms were used in parallel. The number of total presentations was 58, and its details were 8 (General Session), 27 (Student session), 6 (Poster Session), 16 (Organized Session), and 1 special invited speech.

The number of total attendee this year was 123, which is decreased from that of the last year conference, celebrating 50 years of the establishment of IIEEJ, but the conference was held quite lively.

The paper submission category of general/student sessions was divided into Journal Track (JT) and Conference Track, as same as that of the 50th annual conference. To submit the paper to JT Track, authors should prepare the manuscript for conference (2-4 pages) and the manuscript for journal (8 pages) individually. The manuscript for journal will be reviewed with the same process as the usual journal paper. By the way, the special issue of JT papers of IIEEJ Annual Conference 2022 was published on January 2023, and 7 Ordinary Papers and 2 Practice Oriented Papers were contained in it. No English JT papers were received in the IIEEJ Annual Conference 2022.

Concerning the special issue on JT papers in IIEEJ Annual Conference 2023, papers written in Japanese will be published in January 2024 issue, and papers written in English are published in this issue.

The number of the JT paper contained in this issue is one which proposes Modality Former, a deep learning model leveraging multimodality MRI/CT for Hepatocellular Carcinoma (HCC).

In the Annual Conference 2024, we will be continuing to call for JT paper submission, and we would ask the IIEEJ members to submit more papers since this opportunity will be very stimulating to enrich your research results.

Last but not least, we would like to thank all the reviewers and editors for their contribution to improve the quality of papers. We would also like to express our deepest gratitude to the members of the editorial committee of IIEEJ and the staff at the IIEEJ office for various kinds of support.

IIEEJ **Paper** -- *Special Issue on Journal Track Papers in IIEEJ Annual Conference 2023* --

ModalityFormer for Preoperative Early Recurrence Prediction of Hepatocellular Carcinoma Using Multimodality MRI or CT and Clinical Data

Gan ZHAN†, Fang WANG††, Yinhao LI†, Weibin WANG†††,
Qingqing CHEN††, Lanfen LIN††††, Hongjie HU††, Yen-Wei CHEN† (*Member*)

† College of Information Science and Engineering, Ritsumeikan University, Japan,
†† Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, China,
††† Research Center for Healthcare Data Science, Zhejiang Lab, China,
†††† College of Computer Science and Technology, Zhejiang University, Hangzhou, China.

<**Summary**> Hepatocellular carcinoma (HCC), a prevalent liver cancer, poses a substantial mortality risk. Surgical resection is the primary treatment choice, but post-resection recurrence challenges patient outcomes, especially in early recurrenced cases. Developing preoperative early recurrence prediction methods is crucial for personalized treatment plans, improving survival time for HCC patiens. Existing clinical data-based predictions overlook imaging modalities, while radiomics-based methods suffer from limitations imposed by predefined features. In this light, we propose ModalityFormer, a deep learning model leveraging multimodality MRI/CT for HCC early recurrence prediction task. ModalityFormer utilize tranformer architecture to capture inter-modality context and utilize adaptive fusion module to effectively combine prediction logits of all MRI/CT modalities. Consequently, promising results are achieved in our prediction task. Furthermore, we introduce a fusion model called ModalityFormer++, which integrates multimodality MRI/CT with clinical data. Through detailed experiments, we demonstrate that our ModalityFormer model outperforms other state-of-the-art methods. Additionally, ModalityFormer++ exhibits superior performance compared to models relying solely on multimodality MRI/CT or clinical data.

Keywords: Early recurrence prediction, HCC, Multimodality MRI/CT, Transformer, Adaptive fusion module

1. Introduction

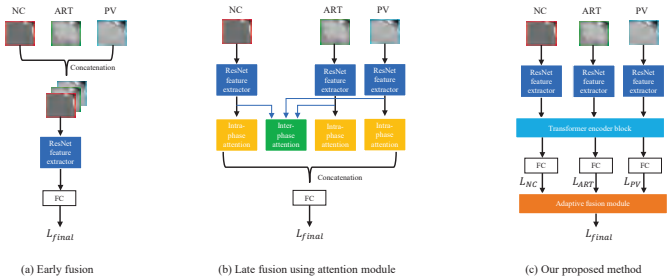
Hepatocellular carcinoma (HCC), a primary liver cancer, poses a significant threat globally, ranking fifth in cancer frequency and second in cancer-related deaths[1]. Predominant in East Asia and sub-Saharan Africa, it constitutes 82% of liver cancer cases[2]. Established HCC treatments include surgical resection, liver transplantation, targeted therapy, immunotherapy, chemoembolization, and radiofrequency ablation. Surgical resection, primarily recommended per clinical guidelines[3)-5)] for HCC patients in the early stage, presents a high risk of recurrence post-resection, exceeding 10% within one year and reaching 70-80% within five years[6]. Identifying patients at high risk of early recurrence is crucial for personalized strategies, ultimately improving survival[7].

Various methods exist for preoperative early recurrence prediction in post-resection HCC patients[8)-12)]. Clinical data methods, using machine learning to map clinical variables for prediction, a limitation in clinical-based methods is: they neglect medical imaging information from MRI and CT scans. These imaging modalities are vital non-invasive tools for detecting HCC malignancy, impacting patient prognosis in terms of recurrence. Subjective clinical lesion assessments are insufficient to replace the role of medical imaging in early recurrence prediction. Radiomics[9), 13)], offering a quantitative approach, extracts robust image features from MRI and CT scans, but it requires tedious feature selection to identify critical features relevant to the task.

In recent years, deep learning has proven highly effective in medical image analysis, surpassing traditional radiomics methods[14] due to automatic extraction for task-relevant features without manual intervention. Approaches have been proposed for early recurrence prediction in HCC using deep learning. Wang et al[10] presented a model using early fusion to combine CT scans for prediction(shown in (a) of **Fig.1**). Wang et al further introduced the joint loss to extract critical modality-specific features[11]. Later, Wang et al proposed the Deep

Phase Attention model[12], incorporating intra-phase attention for modality-specific features and inter-phase attention for inter-modality features (shown in (b) of Fig.1). However, the connections between modalities and their relation to recurrence are not well-explored. Certain studies[15] suggest that the self-attention mechanism in Transformer architecture outperforms conventional attention modules in context modeling.

With these observations, we propose a deep learning method, ModalityFormer, for HCC early recurrence prediction using multimodality MRI or CT. Compared to single-modality MRI/CT, multimodality facilitates the assessment of malignancy severity, a crucial factor in HCC recurrence[16], thus we conduct our study on multimodality MRI/CT. Inspired by the robust contextual modeling capability of Transformer models[17], ModalityFormer formulates each modality image feature as a token of the MRI/CT modality sequence, then it utilizes the Transformer architecture to capture the inter-modality correlations at the feature level, and merges prediction logits of each MRI modality at the decision level by the adaptive fusion module. ModalityFormer is built based on the Transformer based model[18], unlike it, the ModalityFormer first produce the modality-specific prediction logits (shown in (c) of Fig.1), and then it further combines them at the decision level using adaptive fusion module. Although averaging can simply fuse the predictions logits from all modalities, considering the differences among modalities, adaptive fusion module assigns a learnable weight to each prediction logits, and then it conducts the weight sum to obtain the final prediction logits. Based on the ModalityFormer model, we also proposed a fusion model (ModalityFormer++) to combine the multimodality MRI/CT with clinical data. Extensive experimental results reveal that our ModalityFormer outperforms state-of-the-art methods, and our ModalityFormer++ combining multimodality MRI/CT with clinical data surpasses models using only multimodality MRI/CT or clinical data.



(a) Early fusion (b) Late fusion using attention module (c) Our proposed method

Fig. 1 Comparison of our proposed method with existing deep learning methods on our prediction task, the difference lies in exploring inter-modality correlations using Transformer not conventional attention, and fusing multimodality prediction logits not features.

Our study's contributions are as follows: (1) We innovate by treating each MRI/CT modality image feature as a token in the modality sequence, effectively capturing inter-modality correlations using the Transformer architecture. (2) Unlike existing methods[10)-12)] that use concatenation or attention modules for feature-level fusion, we introduce an adaptive fusion module for decision-level fusion of modality prediction logits. (3) Experimental results on multimodality MRI and CT scans show our approach outperforms existing methods for HCC early recurrence prediction.

2. Materials

This study is approved by Sir Run Run Shaw Hospital, Zhejiang University School of Medicine and Ritsumeikan University. The multimodality MRI and aligned clinical data are sourced from Run Run Shaw Hospital. In this retrospective study (2012-2019), 659 HCC patients underwent liver resection and were pathologically confirmed at Sir Run Run Shaw Hospital Affiliated to Medical College of Zhejiang University. All patients had preoperative enhanced MRI; exclusion criteria include: (1) Received prior anti-tumor treatments, such as TACE, RFA; (2) Incomplete clinical data; (3) More than 30 days between preoperative MR examination and surgery; (4) Poor image quality, such as the presence motion and other artifacts; (5) Less than 2 years follow-up after surgery. Our study includes 289 patients, with 108 (37.4%) having early recurrence (ER) within 2 years and 181 (62.6%) identified as non-early recurrence (NER: more than 2 years or no recurrence)[19]. Patients primarily treated with surgical resection were prioritized. Besides MRI, we also test on multimodality CT scans and aligned clinical data. For CT data, detailed information was introduced in existing work[11].

2.1 Magnetic resonance imaging(MRI)

Multimodality MRI (Magnetic Resonance Imaging) offers valuable insights into early recurrence through various imaging modalities. Axial fat-saturated T2-weighted images (T2) provide excellent contrast, delineating tumor size, location, and relationships with surrounding tissues precisely. Diffusion-weighted imaging (DWI) reveals tissue cellularity and tumor edge localization based on wa-

ter molecule diffusion. The arterial phase (T1_ART) of T1-weighted images highlights HCC lesions and vascular invasion, exposing tumor blood supply. Portal vein phase (T1_PV) aids in distinguishing HCC lesions from liver parenchyma through visualization of hepatic portal blood supply[20),21)]. **Figure 2** illustrates these modalities for an NER and an ER patient, showcasing unique textural characteristics of the tumor area. Even for senior radiologists, visually distinguishing between NER and ER patients based on tumor differences is challenging.

Table 1 details the imaging parameters for T1, T2, and DWI. Dynamic contrast-enhanced imaging was acquired at arterial (20 s), portal venous (60 s), and delayed (180 s) phases post-injection of gadopentetate dimeglumine (magnevistt®, Bayer Schering Pharma, Berlin, Germany). The injection rate was 2.5 mL/s, followed by a 15-mL saline flush. Each MRI modality includes tumor annotations with regions of interest (ROIs) outlining lesions drawn by a 9-year-experienced abdominal imaging radiologist. To ensure accuracy, all delineated lesions were reviewed and, if needed, modified by another senior radiologist with 12 years of experience. The combined expertise of both radiologists enhances the credibility of the manually delineated regions.

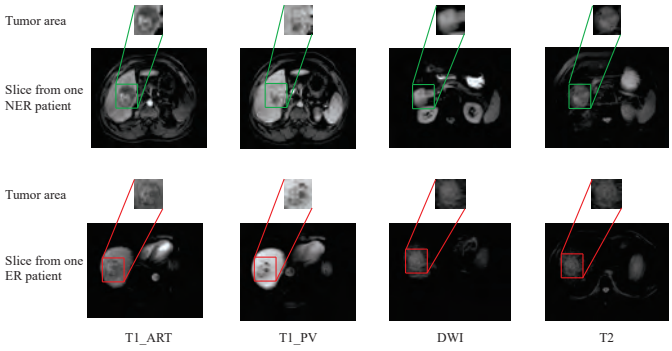


Fig. 2 An example of T1_ART, T1_PV, DWI, and T2 MRI images of NER patients and ER patients in our study.

Table 1 Details of the imaging parameters for T1, T2, and DWI. TR: repetition time; TE: echo time; ST: section thickness; FOV: field of view.(3.0T MRI, GE SIGNA HDxt, Waukesha, Wisconsin, USA)

Parameter	T2-weighted imaging	Dynamic T1-weighted imaging	Diffusion-weighted imaging
TR (ms)	6000	2.57	6000
TE (ms)	90	1.18	60
Flip angle (°)	90	12	90
Matrix	320 × 320	320 × 195	128 × 104
ST (mm)	6	4.4	6
FOV (mm2)	380 × 304	380 × 304	380 × 304

2.2 Clinical data

The clinical data aligned with our multimodality MRI study includes factors like gender (female or male), age, HBV infection (no or yes), liver cirrhosis (no or yes), tumor size, Tumor counts (single or multiple), AFP (alpha-fetoprotein), ALB (albumin), TB (total bilirubin), PLT (platelet count), CP grade (Child-Pugh grade: A or B/C), ALBI (Albumin-Bilirubin Index: 1 or 2/3), and BCLC stage (Barcelona Clinic Liver Cancer stage: 0/A or B/C). Certain pathological factors (as shown in **Table 2**), such as ES (Edmondson-Steiner) grade and microvascular invasion (MVI), weren't included in our study as they're only available post-surgery, unsuitable for preoperative research. We maintain and justify the distinction between the score and individual values. Our study includes a complete set of 13 clinical features, all crucial for predicting early recurrence in HCC. To ensure thorough analysis, we chose not to perform feature selection, incorporating all features into our research.

Table 2 Clinical variables of patients with ER and NER

Clinical factors	Level	Total	NER	ER
Gender	Female/male	51/238	30/151	21/87
Age	≤ 60 / >60	165/124	95/86	70/38
HBV infection	No/Yes	47/242	34/147	13/95
Liver cirrhosis	No/Yes	127/162	86/95	41/67
Tumor size(cm)	≤ 5.0/ >5.0	202/87	141/40	61/47
Tumor counts	Single/Multiple	259/30	169/12	90/18
AFP($\mu g/L$)	≤ 200 / >200	203/86	136/45	67/41
ALB (g/L)	≤ 36 / >36	66/223	32/149	34/74
TB($\mu mol/L$)	≤ 19 / >19	200/89	124/57	76/32
PLT($\times 10^9/L$)	≤ 100 / >100	93/196	54/127	39/69
CP grade	A / B, C	260/29	167/14	93/15
ALBI	1 / 2,3	231/58	147/34	84/24
BCLC stage	0, A / B, C	252/37	170/11	82/26
ES grade	I, II / III, IV	240/49	155/26	85/23
MVI	No/Yes	165/124	117/64	48/60

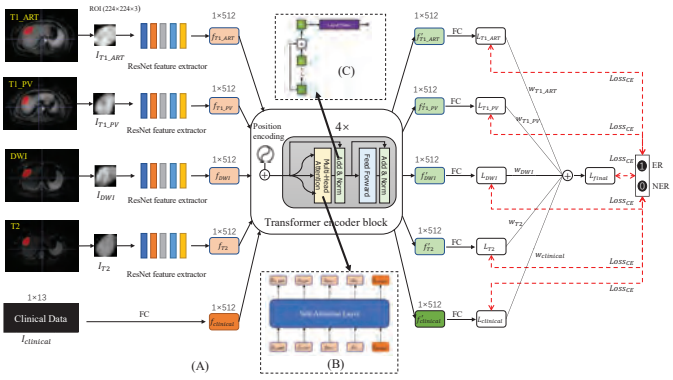


Fig. 3 The overall pipeline of our ModalityFormer model on multimodality MRI and clinical data.

3. Proposed Method

Our proposed prediction method's overall pipeline is in **Fig.3**. Illustrating with the multimodality MRI model: For each patient, we input each MRI ROI into a ResNet-based feature extractor, generating modality-specific image features. Concurrently, clinical data goes through a fully connected (FC) layer for clinical features. MRI modality image features are then combined with clinical features, passing through a Transformer encoder block to capture inter-modality correlations. Next, five FC layers project the output features to prediction logits for each MRI modality and clinical data. Finally, an adaptive fusion module combines all prediction logits, assigning learnable weights for each modality to produce the final prediction logits. In the following section, we would detail the feature extraction module, transformer encoder block, and adaptive fusion module components.

3.1 Feature extraction module

3.1.1 ResNet feature extractor for MRI

ResNet[22] has proven remarkable performance in computer vision domains. We adopt it as the foundation for our modality image feature extractor by removing the last fully connected (FC) layer. Considering our limited dataset size, we choose the compact ResNet18 variant pre-trained on the ImageNet dataset[23),24)]. To adapt the ResNet model, originally trained on natural images with a shape of $224 \times 224 \times 3$, to our MRI data. We selected the most representative slice (containing the largest tumor area), along with its two adjacent slices in each MRI modality. Then we extract the ROI corresponding to the tumor region from these slices. Subsequently, we resized the ROI to $224 \times 224 \times 3$, creating inputs for $I_{T1\_ART}, I_{T1\_PV}, I_{DWI}, I_{T2}$. Each of these inputs was fed into a separate ResNet feature extractor to extract the modality-specific image features.The resulting features in each modality are denoted as f_{T1_ART}, f_{T1_PV}, f_{DWI}, f_{T2}, the shape of them is uniformly 1×512.

3.1.2 FC layer for clinical data

In our study, we include 13 clinical variables ($I_{clinical}$) in our analysis. To effectively use these variables, we use an FC layer, projecting clinical data onto 512-dimensional high-level features. After the FC layer, we apply a ReLU activation function and a dropout layer[25]. The dropout layer helps address overfitting risks by mitigating potential over-parametrization in the clinical branch. Thus, the FC layer enables us to obtain clinical features denoted as $f_{clinical}$.

3.2 Transformer encoder block

The Transformer model[17], initially for natural language processing, uses a self-attention mechanism for inter-token correlations within a sequence. Its versatility extends to computer vision, like the Vision Transformer[26], where images become patches serving as tokens. Inspired by it, we treat each modality image feature and the clinical feature as tokens, totaling five tokens. The intercorrelations within multimodality MRI and between multimodality MRI and clinical data form the token grammar for our prediction task.

Transformer encoder block contains four transformer encoder layers. We concatenate the features above from different modality: $f_{T1\_ART}, f_{T1\_PV}, f_{DWI}, f_{T2}, f_{clinical}$, to obtain the feature vector f_{trans} (shape: 5×512). Through linear projections, we transform the concatenated features into the query ($q\left(f_{trans}\right) = f_{trans}W_q$), key ($k\left(f_{trans}\right) = f_{trans}W_k$), and value ($v\left(f_{trans}\right) = f_{trans}W_v$) vectors in the self-attention mechanism of Eq.(1) (here, we omit the multi-head setting for clarity).

$$Attention(f_{trans}) = softmax(\frac{q(f_{trans})k^T(f_{trans})}{\sqrt{d_k}})v(f_{trans}) \quad (1)$$

Applying the softmax function to the dot product between the query ($q(f_{trans})$) and the transposed key ($k^T(f_{trans})$) could produce the attention map. The attention map is then multiplied with the value ($v(f_{trans})$) vector, encoding the intercorrelations within the tokens. To mitigate the issue of large dot products resulting from the high dimensionality of the $k^T(f_{trans})$ vector (d_k), the dot product result is scaled by $\frac{1}{\sqrt{d_k}}$ to ensure stable gradients during the softmax operation.

Figure 4 illustrates the process of obtaining the T1_ART phase features: By splitting the query representation $q(f_{trans})$ in Eq.(1) along the token dimension, we obtain the query vectors q_{T1_ART}, q_{T1_PV}, q_{DWI}, q_{T2}, $q_{clinical}$ for each modality. Similarly, we obtain the key vectors k_{T1_ART}, k_{T1_PV}, k_{DWI}, k_{T2}, $k_{clinical}$ and the value vectors v_{T1_ART}, v_{T1_PV}, v_{DWI}, v_{T2}, $v_{clinical}$.

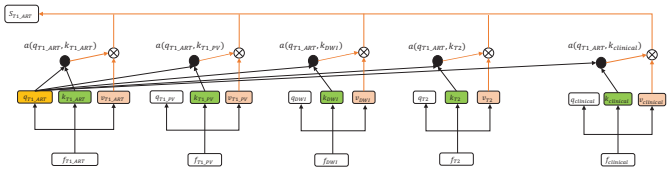


Fig. 4 An example of capturing the intercorrelations between T1_ART phase with other modalities.

Specifically, for the T1_ART phase, the query vector q_{T1_ART} is subject to a normalized dot product operation with the key vectors of all modalities. This operation produces the attention scores, which represent the intercorrelations between the T1_ART phase and all the modalities.Subsequently, these normalized attention scores are multiplied by the value vectors of their respective modalities, effectively extracting the features from the target modality that are associated with the T1_ART phase. Finally, by summing the weighted modality features, we obtain the features that encode the intercorrelations specific to the T1_ART phase, as illustrated in Eq.(2), where $m \in \{T1\_ART, T1\_PV, DWI, T2, clinical\}$.

$$S_{T1\_ART} = \sum a(q_{T1\_ART}, k_m) \times v_m \tag{2}$$

Following the same procedure, we obtain the encoded features for each modality S_{T1_ART}, S_{T1_PV}, S_{DWI}, S_{T2}, $S_{clinical}$ as shown in (B) of Fig.3. And we feed them into the subsequent process (such as (c) in Fig.3).Finally, we obtain the ultimate features for each modality, namely $f\prime_{T1\_ART}$, $f\prime_{T1\_PV}$, $f\prime_{DWI}$, $f\prime_{T2}$, $f\prime_{clinical}$ as shown in Fig.3.

3.3 Adaptive fusion module

Each modality within our MRI data and clinical branch plays a distinct role in our prediction task. Consequently, it becomes imperative to capture their modality-specific features that are pertinent to our task. To address this, we first utilize a FC layer on the ultimate features to generate prediction logits(L_{T1_ART}, L_{T1_PV}, L_{DWI}, L_{T2}, $L_{clinical}$). Subsequently, we apply the sigmoid function to map these logits to prediction probabilities ($y\prime_{T1\_ART}$, $y\prime_{T1\_PV}$, $y\prime_{DWI}$, $y\prime_{T2}$, $y\prime_{clinical}$). By computing the cross-entropy loss between each prediction probability and the corresponding ground truth label (denoted as y) of the patient, we guide our model to learn the modality-specific features. Then, the adaptive fusion module assigns learnable weights (w_{T1_ART}, w_{T1_PV}, w_{DWI}, w_{T2}, $w_{clinical}$) to each prediction logits. The final prediction logits are generated by summing the weighted prediction logits from all modalities as shown in Eq.(3). To get the final prediction probability $y\prime_{final}$, we apply the sigmoid function to the final prediction logits and calculate the cross-entropy loss between the final prediction probability and the corresponding ground truth label of the patient (y). This adaptive fusion mechanism effectively integrates information from different modalities, providing more comprehensive predictions for our task.The loss function used in our proposed method can be summarized as Eq.(4), where $n \in \{T1\_ART, T1\_PV, DWI, T2, clinical, final\}$. Hereby, we perform our HCC early recurrence prediction task with our proposed method.

$$L_{final} = \sum w_m \times L_m \tag{3}$$

$$Loss = \sum Loss_{CE}(y'_n, y) \tag{4}$$

4. Experiments

In the next section, we will introduce the experimental settings, covering the MRI dataset arrangement and employed evaluation metrics. Following that, we'll showcase results for our task using solely multimodality MRI data, and results when incorporating both multimodality MRI/CT and clinical data.

4.1 Dataset arrangement and evaluation metrics

We use Jose Vicente Manjon's preprocessing method[27] to address MRI image artifacts, involving denoising, bias field correction, resampling, and normalization. This pipeline efficiently reduces noise, corrects intensity variations, ensures consistent voxel dimensions, and normalizes intensity values across multimodality MRI data. Manual MRI registration by two radiologists with 9 and 12 years of experience guarantees well-registered ROIs for all modalities. For a fair performance assessment, we conduct a 5-fold cross-validation on our MRI dataset of 289 patients, as outlined in **Table 3**. The evaluation includes metrics such as AUC (Area Under the ROC Curve) and ACC (Accuracy). We calculate average metric values over the 5-fold cross-validation, with our model undergoing 50 epochs of training in each fold. The final trained model is selected based on the checkpoint with the lowest training loss.

4.2 Ablation study

4.2.1 Ablation on single-modality MRI or multimodality MRI

Before delving into multimodality MRI analysis, we justify our research by comparing HCC early recurrence prediction using single-modality and multimodality MRI.

Table 3 Dataset Arrangement of 5-Fold Cross-Validation on multimodality MRI.

Label	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Total
NER	36	29	36	40	40	181
ER	21	29	22	18	18	108
Total	57	58	58	58	58	289

For the single-modality MRI model, we use the ResNet18 architecture. In the case of multimodality MRI prediction, the choice between early fusion and late fusion is debated. Early fusion involves direct concatenation of ROI images, and late fusion employs four separate ResNet models, averaging the prediction logits. Results in **Table 4** show the late fusion model outperforms single-modality MRI models, supporting the importance of researching multimodality MRI for our task. The late fusion model is designated as our baseline due to its superior performance.

4.2.2 Ablation on ModalityFormer

To evaluate our model components, we conducted an ablation study introducing the Transformer encoder block to the late fusion model (Base-Trans), which outperformed the baseline (see **Table 5**). This underscores the transformer encoder block's effectiveness in capturing inter-modality correlations. We further assessed the adaptive fusion module, creating the Base-Ada model, which surpassed the baseline by using the adaptive logits module instead of averaging (see Table 5). Integrating these components into the late fusion model yielded the ModalityFormer model, showing superior performance (see Table 5). Additionally, we introduced ModalityFormer++ by integrating clinical data, outperforming other models (see Table 5). Combining clinical information with image data enhances comprehensive and improved performance in our prediction tasks.

Table 4 Ablation on single-modality MRI or multimodality MRI.

Model	AUC	ACC
T1_ART	0.652	0.640
T1_PV	0.686	0.633
DWI1	0.639	0.609
T2	0.638	0.630
Early fusion	0.645	0.647
Late fusion (baseline)	**0.688**	**0.661**

Table 5 Ablation of ModalityFormer on multimodality MRI and clinical data.

Model	Average-F	Trans-E	Adaptive-F	clinical	AUC	ACC
Base-line	✓				0.688	0.661
Base-Trans	✓	✓			0.693	0.671
Base-Ada			✓		0.691	0.682
Modality-Former		✓	✓		0.713	0.685
Clinical-FC				✓	0.681	0.654
Modality-Former++		✓	✓	✓	0.732	0.730

4.3 Comparison with existing methods

In comparing our ModalityFormer model to existing approaches, including radiomics-based and deep learning-based methods (PhaseNet and DPANet) for multimodality MRI (see **Table 6**), ModalityFormer demonstrated superior predictive performance in HCC early recurrence. Evaluating ModalityFormer++ against state-of-the-art methods (Phase-F, Phase-F-joint, DPA-F, and clinical-KNN) using multimodality MRI and clinical data, ModalityFormer++ achieved better performance (see Table 6). This underscores the model's effectiveness in leveraging both multimodality MRI and clinical data, resulting in improved prediction accuracy. Moreover, applying our method to multimodality CT scans and comparing it with existing methods[10)-12)] (see **Table 7**), our method demonstrated superior results, whether using CT scans alone or combining them with aligned clinical data, further validating its effectiveness.

5. Conclusion

This paper presents ModalityFormer for preoperative early recurrence prediction in HCC using multimodality MRI or CT. The model utilizes separate ResNet feature extractors, treats each modality image feature as a token, employs a transformer encoder block for inter-modality correlations, and combines prediction logits

Table 6 Comparison with existing methods on multimodality MRI and clinical data.

Model	Multi-modality MRI	Clinical data for MRI	AUC	ACC
Radiomics[9)]	✓		0.686	0.636
PhaseNet[10)]	✓		0.645	0.664
DPANet[12)]	✓		0.682	0.671
ModalityFormer	✓		**0.713**	**0.685**
Clinical-KNN[8)]		✓	0.686	0.668
Phase-F[10)]	✓	✓	0.687	0.675
Phase-F-joint[11)]	✓	✓	0.691	0.675
DPA-F[12)]	✓	✓	0.707	0.675
ModalityFormer++	✓	✓	**0.732**	**0.730**

Table 7 Comparison with existing methods on multimodality CT and clinical data.

Model	Multi-modality CT	Clinical data for CT	AUC	ACC
PhaseNet[10)]	✓		0.723	0.695
DPANet[12)]	✓		0.747	0.707
ModalityFormer	✓		**0.781**	**0.710**
Phase-F[10)]	✓	✓	0.825	0.787
Phase-F-joint[11)]	✓	✓	0.833	0.805
DPA-F[12)]	✓	✓	0.869	0.812
ModalityFormer++	✓	✓	**0.878**	**0.826**

through an adaptive fusion module. Additionally, ModalityFormer++ extends the model to incorporate multi-modality MRI/CT and clinical data, exploring intercorrelations for a comprehensive understanding. Experimental results demonstrate the efficacy of both models, with ModalityFormer and ModalityFormer++ outperforming existing approaches, including radiomics-based and deep learning-based methods. Notably, ModalityFormer++ achieves superior performance by integrating both image and clinical modalities, surpassing models relying solely on individual modalities.

Acknowledgments

This work was supported in part by the Grant in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant Nos. 20KK0234, 21H03470, and 20K21821, and in part by the Natural Science Foundation of Zhejiang Province (LZ22F020012), in part by Major Scientific Research Project of Zhejiang Lab (2020ND8AD01), and in part by the National Natural Science Foundation of China (82071988), the Key Research and Development Program of Zhejiang Province (2019C03064), the Program Co-sponsored by Province and Ministry (No. WKJ-ZJ-1926) and the Special Fund for Basic Scientific Research Business Expenses of Zhejiang University (No. 2021FZZX003-02-17).

References

1) K.M.Elsayes, A.Z.Kielar, M.M.Agrons, et al: "Liver Imaging Reporting and Data System: an expert consensus statement," Journal of hepatocellular carcinoma, pp. 29-39 (2017).
2) R.X.Zhu, W.K.Seto, C.L.Lai, M.F.Yuen: "Epidemiology of hepatocellular carcinoma in the Asia-Pacific region," Gut and Liver, Vol.10, No.3, pp. 332-339 (2016).
3) M.B.Thomas, A.X.Zhu: "Hepatocellular Carcinoma: The Need for Progress," Journal of Clinical Oncology, Vol.23, No.13, pp. 2892-2899 (2005).
4) European Association for the Study of the Liver: "EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma," Journal of Hepatology, Vol.69, No.1, pp. 182-236 (2018).
5) J.A.Marrero, L.M.Kulik, C.B.Sirlin, et al: "Diagnosis, Staging and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases," Hepatology, Vol.68, No.2, pp. 723-750 (2018).
6) F.Bray, J.Ferlay, I.Soerjomataram, R.L.Siegel, L.A.Torre, A.Jemal: "Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA: A Cancer Journal for Clinicians, Vol.68, No.6, pp. 394-424 (2018).
7) Z.Cheng, P.Yang, S.Qu, J.Zhou, J.Yang, X.Yang, Y.Xia, J.Li, K.Wang, Z.Yan, D.Wu, B.Zhang, N.Hüser, F.Shen: "Risk factors and management for early and late intrahepatic recurrence of solitary hepatocellular carcinoma after curative resection," HPB, Vol.17, pp. 422-427 (2015).
8) C.Liu, H.Yang, Y.Feng, et al: "A K-nearest Neighbor Model to Predict Early Recurrence of Hepatocellular Carcinoma After Resection," Journal of Clinical and Translational Hepatology, Vol.10, No.4, pp. 600-607 (2022).
9) Y.Zhao, J.Wu, Q.Zhang, Z.Hua, et al: "Radiomics Analysis Based on Multiparametric MRI for Predicting Early Recurrence in Hepatocellular Carcinoma After Partial Hepatectomy," Journal of Magnetic Resonance Imaging, Vol.53, pp. 1066-1079 (2021).
10) W.Wang, Q.Chen, Y.Iwamoto, X.Han, Q.Zhang, H.Hu, L.Lin, Y.Chen: "Deep learning-based radiomics models for early recurrence prediction of hepatocellular carcinoma with multi-phase CT images and clinical data," International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), PP. 4881-4884 (2019).
11) W.Wang, Q.Chen, Y.Iwamoto, P.Aonpong, L.Lin, H.Hu, Q.Zhang, Y.Chen: "Deep fusion models of multi-phase CT and selected clinical data for preoperative prediction of early recurrence in hepatocellular carcinoma," IEEE Access (2020).
12) W.Wang, F.Wang, Q.Chen, S.Ouyang, Y.Iwamoto, X.Han, L.Lin, H.Hu, R.Tong, Y.Chen: "Phase Attention Model for Prediction of Early Recurrence of Hepatocellular Carcinoma with Multi-phase CT Images and Clinical data," Frontiers in Radiology, Vol.8 (2022).
13) R.J.Gillies, P.E.Kinahan, H.Hricak: "Radiomics: images are more than pictures, they are data," Radiology, Vol.278, pp. 563-577 (2015).
14) G.Litjens, T.Kooi, B.E.Bejnordi, et al: "A survey on deep learning in medical image analysis," Medical Image Analysis, Vol.42, pp. 60-88 (2017).
15) M.Guo, T.Xu, J.Liu, Z.Liu, P.Jiang, T.Mu, S.Zhang, R.R.Martin, M.Cheng, S.Hu :"Attention mechanisms in computer vision: A survey," Computational visual media, pp. 331-368(2022).
16) M.Armbruster, M.Guba, J.Andrassy, M.Rentsch, V.Schwarze, J.Rübenthaler, T.Knösel, J.Ricke, H.Kramer: "Measuring HCC Tumor Size in MRI—The Sequence Matters!" Diagnostics, Vol.11, No.11 (2002).
17) A.Vaswani, N.Shazeer, N.Parmar, J.Uszkoreit, L.Jones, A.N.Gomez, L.Kaiser, I.Polosukhin: "Attention is all you need," Advances in neural information processing systems, Vol.30 (2017).
18) G.Zhan, F.Wang, W.Wang, Y.Li, Q.Chen, H.Hu, Y.Chen: "A Transformer-Based Model for Preoperative Early Recurrence Prediction of Hepatocellular Carcinoma with Multi-modality MRI," The Fourth IEEE International Workshop on Machine Learning and Computing for Visual Semantic Analysis, pp. 185-195 (2023).
19) H.Xing, W.G.Zhang, M.Cescon, L.Liang, C.Li, M.Wang, H.Wu, W.Y.Lau, Y.H.Zhou, W.M.Gu, H.Wang, T.Chen, Y.Zeng, M.Schwartz, T.M.Pawlik, M.Serenari, F.Shen, M.C.Wu, T.Yang: "Defining and predicting early recurrence after liver resection of hepatocellular carcinoma: a multi-institutional study," HPB, Vol.22, No.5, pp. 677-689 (2020).
20) Y.Lee, W.H.Jee, Y.S.Whang, C.K.Jung, Y.G.Chung, S.Y.Lee: "Benign versus malignant soft-tissue tumors: differentiation with 3T magnetic resonance image textural analysis including diffusion-weighted imaging," Investigative Magnetic Resonance Imaging, Vol.25, No.2, pp.118-128 (2021).
21) E.Chartampilas, V.Rafailidis, V.Georgopoulou, G.Kalarakis, A.Hatzidakis, P.Prassopoulos: "Current Imaging Diagnosis of

Hepatocellular Carcinoma," Cancers, Vol.14, No.16, pp. 3997 (2022).
22) K.He, X.Zhang, S.Ren, J.Sun: "Deep residual learning for image recognition," Proc. of IEEE conference on computer vision and pattern recognition, pp. 770-778 (2016).
23) J.Deng, W.Dong, R.Socher, J.Li; L.Kai; F.Li: "Imagenet: A large-scale hierarchical image database", Proc. of IEEE conference on computer vision and pattern recognition, pp. 248-255 (2009).
24) A.Kolesnikov, L.Beyer, X.Zhai, J.Puigcerver, J.Yung, S.Gelly, N.Houlsby: "Big Transfer (BiT): General visual representation learning," European conference on computer vision, pp. 491-507 (2020).
25) N.Srivastava, G.Hinton, A.Krizhevsky, I.Sutskever, R.Salakhutdinov;: "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," The journal of machine learning research, pp. 1929-1958 (2014).
26) A.Dosovitskiy, L.Beyer, A.Kolesnikov, et al: "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint arXiv:2010.11929 (2020).
27) J.V.Manjón: "MRI Preprocessing," Imaging Biomarkers, pp. 53-63 (2016).

(Received June 26, 2023)
(Revised November 15, 2023)

Gan Zhan
He received bachelor's degree and Master's degree from Northeast University, Shenyang, China in 2017 and 2020, respectively. He is now pursuing PhD at Ritsumeikan University, Kusatsu, Japan. His research interests include computer vision, medical image analysis, and deep learning.

Fang Wang
She received the bachelor of medicine degree from the School of Medicine, China Three Gorges University, in 2019. Before graduation, she had a one-year internship in major medicine to learn clinical knowledge. She is currently pursuing the Ph.D. degree in medicine with Zhejiang University School of Medicine, China.

Yinhao Li
He received the B.E. degree from Southeast University Chengxian College, Nanjing, China, in 2013, and the M.E. and D.E. degrees from Ritsumeikan University, Kusatsu, Japan, in 2018 and 2021, respectively. He is currently an Assistant Professor with the College of Information Science and Engineering, Ritsumeikan University, Japan. His research interests include computer vision, medical image analysis, and deep learning.

Weibin Wang
He was born in Guangxi Province, China, in 1994. He received a bachelor's degree in 2017 from Northeast University, Shenyang, China. He received his Master's degree in 2019, and his Doctor's degree in 2022, both from Ritsumeikan University, Shiga, Japan. In 2023, he joined the Zhejiang Laboratory as a postdoctoral researcher, specializing in the field of intelligent healthcare big data.

Qingqing Chen
She received the doctor's degree in medicine from Zhejiang University in 2021. Currently, she is working at Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, China. Her research mainly focuses on the artificial intelligence of liver imaging, such as lesion classification, LI-RADS category, and prediction or prognosis of HCC.

Lanfen Lin
She (Member, IEEE) received the BS degree from the Northwest University of Technology, in 1990, and the PhD degree from the Northwest University of Technology, in 1995. She is a professor of the College of Computer Science and Technology, Zhejiang University. Her research interests include medical image analysis, recommender system, and data mining.

Hongjie Hu
He received the master's and doctor's degrees in medicine from Zhejiang Medical University in 1992 and 1998, respectively. He has been the Chief Physician and a Doctoral Supervisor with the Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, since 2015. He has engaged in medical imaging for over 30 years. He went to some place for further education at Loma Linda University Medical Center in 2000, Myo Clinic in 2008, and Cleveland University Clinic, USA. He has rich clinical experience in cardiothoracic diagnosis and interventional radiotherapy.

Yen-Wei Chen (*Member*)
He received his B.E. degree in 1985 from Kobe University, Kobe, Japan. He received his M.E. degree in 1987, and his D.E. degree in 1990, both from Osaka Univ., Osaka, Japan. He is currently a professor at the college of Information Science and Engineering, Ritsumeikan University, Japan. His research interests include computer vision, medical image analysis.

IIEEJ **Short Paper**

Image Quality Evaluation of Computer-Generated Hologram to Compare Various Calculation Methods

Hiroshi YOSHIKAWA† (*Member*), Takeshi YAMAGUCHI†

† Dept. Computer Engineering, Nihon University

<**Summary**> Image quality of the computer-generated holograms are usually evaluated subjectively. For example, the reconstructed image from the hologram is compared with other holograms or the original image. In previous research, authors proposed image quality evaluation by peak signal-to-noise ratio and diffraction efficiency. In the present research, the structural similarity index (SSIM) is evaluated which is considered to have good agreement to perceptual image quality. Theory and numerical experimental results are shown on Fourier transform transmission hologram of both amplitude and phase modulation. From numerical simulations, bipolar intensity method, or real number calculation gives better quality than that of classical complex number calculation. Phase hologram is much brighter than amplitude hologram, however SSIM becomes lower except Kinoform that has optimized phase in the image plane.

Keywords: computer-generated hologram, image assessment, image quality control, diffraction efficiency, structural similarity

1. Introduction

Computer-generated hologram (CGH) is widely used for three dimensional displays[1)–4)]. Its image quality is usually evaluated subjectively, or compared with the other holograms. For example, the quality of the information reduced hologram is evaluated subjectively by the double-stimulus impairment scale method to compare with the original image[5)]. As the objective quality assessment, the signal-to-noise ratio is evaluated to improve image quality of the binary phase hologram[6)]. To evaluate three dimensional signal-to-noise ratio, VSNR (volume signal-to-noise ratio) is proposed[7)]. Other indices of CGH, such as contrast ratio, cross talk, color dispersion, and uniformity were evaluated[8)].

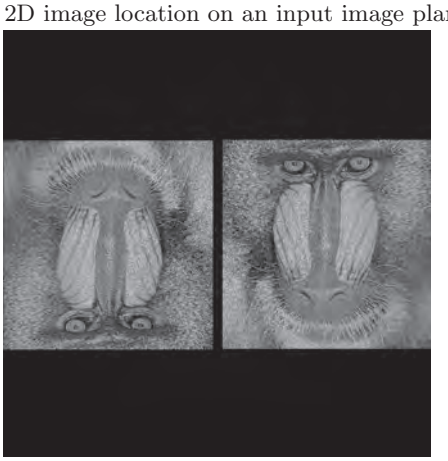
In our previous researches, the quality is evaluated by both the brightness and the peak signal-to-noise ratio (PSNR) of the reconstructed image with various calculation methods and parameters on both the amplitude and phase modulation of the Fourier transform transmission hologram[9)–11)]. In this paper, new insights have been obtained by analyzing the image quality of amplitude and phase holograms in more detail such as extension of parameter range. Also, quality measurement metrics are changed from PSNR to Structural Similarity Index (SSIM)[13)]. One reason is that in 2D image quality evaluation, it is known that the PSNR does not always reflect visual perception. Since the human visual perception is highly adopted for extracting structural information, Structural Similarity Index (SSIM) is proposed as an objective method for assessing perceptual image quality. Another reason is that SSIM is more sensitive to grainy noise than PSNR. Because the grainy noise is major noise source in hologram reconstruction as appeared in **Fig. 1**(b) and the Figure referred in **3.3**.

2. Calculation and Evaluation Theory

2.1 Classical calculation of computer-generated Fourier hologram

The Fourier hologram can be calculated with the Fourier transform of an original image[1)]. Figure 1(a) shows the image location in the input image plane for the Fourier transform and Fig. 1(b) shows a numerically reconstructed image from the hologram. In the present paper, 2D input image size of 500^2 pixel and the hologram size of $1,024^2$ pixel is used. As one can see from Fig. 1(b), the reconstructed image includes the desired image appeared as same position of the original image and the conjugate image that appears as the point symmetry to the center. The reconstructed image also includes grainy noise that reduces image quality. The direct light (or non-diffracted light) is eliminated numerically in Fig. 1(b), by subtract the average value of the inter-

(a) 2D image location on an input image plane



(b) A simulated reconstruction image from the Fourier hologram

Fig. 1 Image location and reconstructed image of the Fourier transform hologram

ference fringes from the interference fringes. It usually appear at the center of the reconstructed image. In all reconstruction results except Fig. 1(b), the direct light is taken into account to evaluate image quality. The original image should be placed off-center not to overlap with the direct light and the conjugate image. In this paper, the original image is located center in vertical and right most side in horizontal. For the hologram calculation, the pixel value other than the original 2D image is set to zero. The random phase is multiplied to each pixel to make the reconstructed image diffusing and bright[1]. Then 2D Fourier transform is applied to the amplitude distribution of $o(x, y)$ with the random phase on the input image plane, and the result of $O(X, Y)$ represents the complex amplitude of the object beam on the hologram plane. If the reference beam is collimated and its direction is perpendicular to the hologram, the complex amplitude of the reference beam $R(X, Y)$ can be represented as the real-valued constant r. The total complex amplitude on the hologram plane is the interference of the object and reference beam, represented as $O(X, Y) + r$. The total intensity pattern,

$$\begin{aligned} I(X,Y) &= |O(X,Y) + r|^2 \\ &= |O(X,Y)|^2 + r^2 + rO(X,Y) + rO^*(X,Y) \\ &= |O(X,Y)|^2 + r^2 + 2r\Re\{O(X,Y)\}, \end{aligned} \tag{1}$$

is a real physical light distribution on the hologram. Where, $\Re\{C\}$ takes the real part of the complex number C, and C^* means the conjugate of C. At the right most hand of the Eq. (1), the first term represents the object self-interference and the second is the reference beam intensity. The third term is the interference of the object and the reference beams, and contains holographic information. In the numerical experiments, a scaled fringe intensity pattern $I_s(X, Y)$ is used defined as:

$$I_s(X,Y) = \frac{I(X,Y)}{I_{\max}}, \tag{2}$$

where $I_{\max}$ is the maximum value of $I(X, Y)$. The scaled intensity takes the value between 0 to 1.

2.2 Normalization to keep the diffraction efficiency of the hologram

In the optical hologram, it is well known that the larger beam ratio, the intensity ratio of the reference beam over the object beam, makes the noise on the reconstructed image smaller. It, however, also reduces the diffraction efficiency or the intensity of the reconstructed image, because the higher beam ratio reduces the fringe visibility or the interference modulation depth[15]. In the CGH, it is easy to keep the fringe visibility large by simply normalizing the hologram intensity instead of Eq. (2) defined as:

$$I_n(X,Y) = \frac{I(X,Y) - I_{\min}}{I_{\max} - I_{\min}}, \tag{3}$$

where $I_{\min}$ is the minimum value of $I(X, Y)$.

2.3 Calculation without the object self-interference

In the CGH, it is quite easy to use only the interference term $I_b(X, Y) = 2r\Re\{O(X, Y)\}$ of Eq. (1). This idea is proposed at very early stage of CGH research[2] . Since the interference term I_b takes both positive and negative values, it is also called "bipolar intensity[16]." The normalization defined in Eq. (3) is also applied to make final fringe intensity I_n positive. "Bipolar intensity" seems quite similar to normalized hologram, but it does not contain the object self-interference that is major source of noise. In the normalized hologram, increasing beam ratio defined in **2.7** can reduce the contribution

of the object self-interference. Therefore, "bipolar intensity" can be considered as equivalent to set the beam ratio of normalize hologram to infinity. However, calculation is simpler than that of normalized one.

2.4 Numerical reconstruction

Numerical reconstruction from the calculated hologram is performed by the inverse Fourier transform. If the 0-th order diffraction, or non-diffracted light is very strong, it masks the 1-st order diffraction, or the reconstructed image. Therefore, to show the reconstructed image clearly, the 0-th order diffraction light is removed numerically as shown in Fig. 1(b). However, the intensity of the 0-th order diffraction light must be taken into account to calculate the diffraction efficiency.

For numerical reconstruction of the transmission amplitude hologram, the amplitude transmittance $t(X,Y)$ of the CGH is assumed to be equal to the fringe intensity $I(X,Y)$ as[15]:

$$t(X,Y) = I_n(X,Y). \tag{4}$$

Then, $t(X,Y)$ is inverse Fourier transformed to obtain the reconstructed image.

For the transmission phase CGH, the complex amplitude transmittance $t(X,Y)$ is assumed as:

$$t(X,Y) = \exp[-i2\Delta\phi I(X,Y)]. \tag{5}$$

Then, $t(X,Y)$ is inverse Fourier transformed to obtain the reconstructed image. In the case of a sine-wave phase grating, the maximum diffraction efficiency of 33.8 % is obtained at $\Delta\phi = 0.59\pi$[15]. Therefore, this value of $\Delta\phi$ is used unless otherwise denoted.

2.5 Kinoform

From previous researches[9),10)], it is found that the image quality of both amplitude and phase transmission holograms depend on the random phase, applied on the original image as a diffuser. Therefore, the image quality of Kinoform[17)]is evaluated as a CGH with the optimized diffuser[11)]. Since the original Kinoform also uses a random phase, Gerchberg and Saxton algorithm[18)]is applied to obtain the optimized diffuser. The formulation and procedure described in **8.3** of Reference 19) is utilized to calculate the Kinoform, which is summarized as:

1. Prepare the original image whose pixel size is same as the Kinoform, and apply random phase to the image. Let $o(x,y)$ as complex amplitude in the image plane.
2. Fouriere transform $o(x,y)$ to obtain $O(X,Y)$ in the Kinoform plane. Phase component of $O(X,Y)$ is the Kinoform.
3. Replace amplitude component of $O(X,Y)$ to one while keeping phase component.
4. Inverse Fouriere transform $O(X,Y)$ to obtain $o(x,y)$ in the image plane. Amplitude component of $o(x,y)$ represents reconstructed image and used to evaluate quality.
5. Replace amplitude component of $o(x,y)$ to the original image while keeping phase and go to procedure 2.

Procedures 2-5 are repeated until the Kinoform satisfies the requirement.

2.6 Diffraction efficiency

The diffraction efficiency (DE) is defined as the ratio of the intensities of the reconstructed image and the illumination light. It gives the brightness of the reconstructed image. In the numerical experiments, the reconstructed image intensity is obtained by summing up all intensities in the reconstructed image area as same size and position of the original image in the input image plane.

2.7 Beam ratio

The beam ratio (BR) is defined as the intensity ratio of the reference and the object beam, written as:

$$\mathrm{BR} = \frac{|R|^2}{|O|^2}. \tag{6}$$

2.8 Structural Similarity Index (SSIM)

In the present paper, the reconstructed image is extracted from reconstructed image plane such as Fig. 1(b) and normalized to 8-bit gray-scale image that has same mean intensity of the original image, then the SSIM[13)]is calculated from the luminance, contrast and structure as,

$$\mathrm{SSIM}(x,y) = \frac{(\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{7}$$

where x and y indicate original and reconstructed images, μ is average luminance, σ is dispersion and c_1 and c_2 are constants. The SSIM is evaluated in the part of images, for example 11 $\times$ 11 pixel block with certain weighting function. To evaluate entire image, a mean SSIM is used. SSIM takes value from 0 to 1. The value of 1 means tow images are identical, and the value of 0 means two images are quite different. From experimental results shown in original SSIM paper[13)], the SSIM value over 0.9 seems quality image. Although the original SSIM depends on the intensity difference of images, the intensity of the reconstructed image is adjusted to have the same mean value of the original image in the present research.

3. Results of Numerical Experiments

An image used in the present research is Mandrill from the USC-SIPI image database[20], gray scaled and re-sized to fit the hologram. The pixel number of the hologram is denoted as $N \times N$, and N is set to 1,024. The width and the height of the original 2D image is denoted as W and H in pixel, respectively, and $W = H = 500$. For Kinoform, hologram and 2D image are set to the same size as $N = W = H = 512$. In the previous research[14], it is found that the diffraction efficiency value depends on the characteristics of the random phase multiplied to input image. For example, in the classical type of amplitude hologram with beam ratio of 10, average diffraction efficiency is 0.56 %, ranging from 0.38 % to 0.71 % and standard deviation of 0.06 %. Therefore, each result is obtained as the average value of at least 20 different random phases. In contrast, SSIM is stable from the random phase. Therefore, only one calculation is enough for Kinoform, because Kinoform always has 100 % of diffraction efficiency.

3.1 Evaluation results for the amplitude hologram

Figure 2 shows DE and SSIM of the classical calculation (solid lines) and the normalized hologram (dashed lines) against the beam ratio. As the beam ratio increases, SSIM increases. This is because the object self-interference remains large when the beam ratio is small and it affect as noise to reduce SSIM. While DE of the classical calculation decreases as the beam ratio increases, the DE for the normalized hologram remains almost constant. Therefore, the normalized hologram could obtain higher SSIM while keeping the brightness. In the classical calculation, SSIM takes the peak value of 0.966 around BR of 1,000 then decrease. This is caused by the quantization error because the CGH is represented as eight bit unsigned integer to fit physical device such as LCOS.

The bipolar intensity hologram, or the hologram without the Object Self-Interference is evaluated. The beam ratio does not make sense for this hologram. It gives SSIM of 0.997 and DE of 0.56 %. The DE and SSIM for three transmission amplitude holograms are listed in **Table 1**. It can be said that the bipolar intensity hologram gives the best result among three hologram calculation methods.

3.2 Evaluation results for the phase hologram

Figure 3 shows the DE and SSIM of the classical calculation (solid lines) and normalized hologram (dashed lines) against the beam ratio. The DE of the phase hologram becomes over ten times larger than that of the amplitude hologram. As same as the transmission amplitude hologram shown in Fig. 2, as the beam ratio increases, SSIM increases. For the normalized hologram, although the DE remains almost constant, the SSIM does not increase as the classical hologram does. In the classical calculation, SSIM takes the peak value of 0.964 around BR of 1,000 then decrease. The reason is quantization error as same as the amplitude hologram.

The bipolar intensity phase hologram gives SSIM of 0.870 and DE of 6.91 %, those are similar to the normalized phase hologram. As mentioned in **2.4**, the phase modulation amplitude $\Delta\phi = 0.59\pi$ is used because it gives the best DE for the sine-wave grating. However, the hologram in the present paper is calculated from the 2D image and there is no guarantee that the value of 0.59π gives the best quality. Therefore, DE and SSIM against $\Delta\phi$ are analyzed for the bipolar hologram and the result is shown in **Fig. 4**. When $\Delta\phi$ increases, DE is also increases. However, SSIM is going down and this means increasing $\Delta\phi$ is not practical. In contrast, as $\Delta\phi$

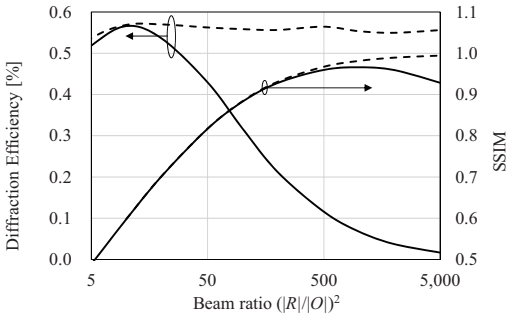


Fig. 2 Diffraction efficiency and SSIM against beam ratio for transmission amplitude holograms

Table 1 Comparing DE and SSIM for various calculations of the transmission amplitude hologram

type	DE [%]	SSIM	beam ratio
classical	0.07	0.966	1,000
normalized	0.56	0.994	5,000
bipolar	0.56	0.997	-

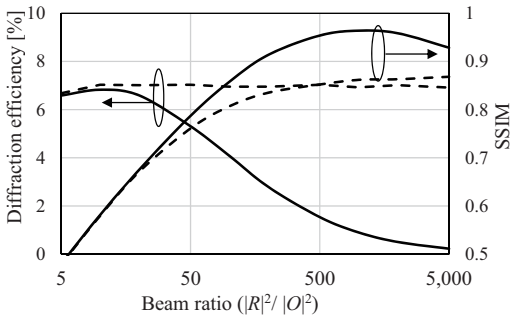


Fig. 3 Diffraction efficiency and SSIM against beam ratio for transmission phase holograms

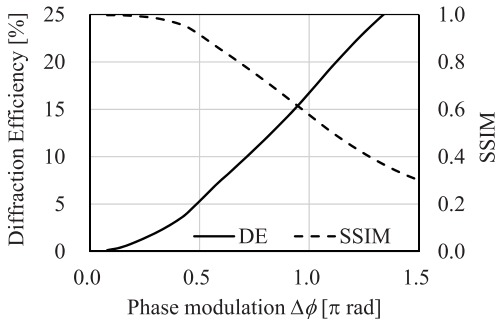


Fig. 4 Diffraction efficiency and SSIM against phase modulation amplitude for transmission phase holograms

Table 2 Comparing DE and SSIM for various calculations of the transmission phase hologram

type	DE [%]	SSIM	$\Delta\phi$ [π rad]	beam ratio
classical	0.90	0.964	0.59	1,000
normalized	6.92	0.868	0.59	5,000
bipolar	6.91	0.870	0.59	-
bipolar2	3.15	0.966	0.38	-

decreases, the DE also decreases but the SSIM increases. For example, when $\Delta\phi = 0.38\pi$, DE decreases to 3.15 %. However, SSIM reaches 0.966, as good as classical calculation with beam ratio of 1,000, but diffraction efficiency is 3.5 times higher. It means that the image quality of the transmission phase hologram can be controlled by changing the phase modulation amplitude.

The DE and SSIM for various transmission phase holograms are summarized in **Table 2**. Among phase holograms, the 'bipolar2' shows the best balanced values of DE and SSIM. As shown in the Table, the phase hologram provides more than ten times larger DE than that of the amplitude hologram. However, SSIM is not as good as the amplitude hologram. Therefore, it can be said that the reconstructed image from the phase hologram is generally brighter but noisy to compare with the amplitude hologram.

3.3 Evaluation of the Kinoform

Image quality of the Kinoform[17]is evaluated to investigate the performance of CGH with the optimized diffuser instead of the random phase. 2D image and Kinoform size are both 512^2 pixel. The SSIM against the number of iterations are plotted in **Fig. 5**. Since the reconstructed image size is same as the reconstruction image plane and assuming no absorption in the CGH, the DE always takes the value of 100 %. As shown in the figure, the SSIM increases continuously as the number of iteration increases. **Figure 6** shows some examples of the reconstructed image from the Kinoform. Figure 6(a) is the reconstructed

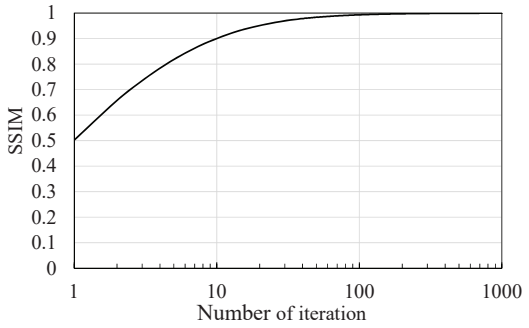
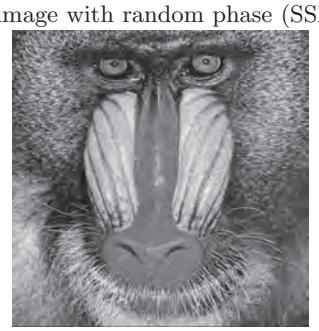


Fig. 5 SSIM against number of iteration for Kinoform

(a) Initial image with random phase (SSIM = 0.50)



(b) Image after 10 iterations (SSIM = 0.90)

(c) Image after 100 iterations (SSIM = 0.99)

Fig. 6 Reconstructed images from Kinoform

image from the initial Kinoform with random phase and SSIM is as low as 0.50. After 10 iterations, SSIM is improved as 0.90 as shown in Fig. 6(b). Figure 6(c) shows the reconstructed image of 100 iterations

The image quality of the transmission phase hologram may be also improved with optimized diffuser obtained by such as the iterative Fourier transform algorithm[18].

4. Conclusion

Image quality of CGH is evaluated objectively with the diffraction efficiency (DE) and Structural Similarity In-

dex (SSIM). Among several computation methods, bipolar intensity method shows the best quality image. This was known empirically, but was supported by objective evaluation. In comparison between the amplitude hologram and the phase hologram, it is found that the phase hologram gives more than 10 times brighter but noisy image. However, the phase modulation is assumed proportional to the fringe intensity. Therefore, other type phase holograms should be investigated. As a special type phase hologram, Kinoform is evaluated. Although the initial Kinoform with random phase does not show the quality image, optimized phase with iteration method provides quality image with 100 % diffraction efficiency.

Although both SSIM and the DE are used to evaluate the image quality subjectively, it is required to investigate the balance of these two values. For example, if the one hologram has higher SSIM but lower DE, and the other has lower SSIM but higher DE, how can we decide which hologram is better? Since current results only evaluate very simple Fourier holograms, evaluations for 3D image is desired for the future work.

Acknowledgements

This work was supported by JSPS KAKENHI grant number JP16K00282.

References

1) A. W. Lohmann, D. P. Paris: "Binary Fraunhofer Holograms, Generated by Computer", Applied Optics, Vol. 6, No. 10, pp. 1739–1748 (Oct. 1967).
2) J. P. Waters: "Holographic image synthesis utilizing theoretical methods", Applied Physics Letters, Vol. 9, No. 11, pp. 405–407 (Dec. 1966).
3) T. Huang: "Digital holography", Proc. of the IEEE, Vol. 59, No. 9, pp. 1335–1346 (Sep. 1971).
4) H. Yoshikawa: "Computer-Generated Holograms for White Light Reconstruction", in Digital Holography and ThreeDimensional Display, Poon, T.-C., ed., ch. 8, pp. 235–255, Springer (2006).
5) H. Yoshikawa, K. Sasaki: "Image scaling for electroholographic display", Proc. of SPIE, Vol. 2176, pp. 12–22 (May 1994).
6) E. Buckley: "Real-Time Error Diffusion for Signal-to-Noise Ratio Improvement in a Holographic Projection System", Journal of Display Technology, Vol. 7, No. 2, pp. 70–76 (Feb. 2011).
7) K. Suzuki, Y. Sakamoto: "Measurement method for objective evaluation of reconstructed image quality in CGH", Proc. of SPIE, Vol. 8644, pp. 864412(1-7), (Mar. 2013).
8) W. Seo, H. Song, J. An, J. Seo, G. Sung, Y.-T. Kim, C.-S. Choi, S. Kim, H. Kim, Y. Kim, Y. Kim, Y. Kim, H.-S. Lee, S. Hwang: "Image Quality Assessment for Holographic Display", Proc. of Electronic Imaging, Vol. 29, pp. 186–190 (Jan. 2017).
9) H. Yoshikawa, T. Yamaguchi: "Image Quality Evaluation of a Computer-Generated Hologram", Digital Holography & 3-D Imaging Meeting, paper DT2A.8, OSA, Shanghai, China (May 2015).
10) H. Yoshikawa, T. Yamaguchi: "Image Quality Evaluation of a Computer-Generated Phase Hologram", 10th International Symposium on Display Holography, paper 4–4 (July 2015).
11) H. Yoshikawa, T. Yamaguchi, H. Uetake: "Image quality evaluation and control of computer-generated holograms", Proc. of SPIE, Vol. 9771, pp. 97710N(1-9) (Mar 2016).
12) Y. Nakaguchi, T. Yamaguchi, H. Yoshikawa: "Image quality improvement of large size image type computer-generated hologram", The 2012 International Workshop on Advanced Image Technology, pp. 572–575 (Jan. 2012).
13) Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli: "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Transactions on Image Processing, Vol. 13, No. 4, pp. 600–612 (Apr. 2004).
14) H. Uetake, H. Yoshikawa, T. Yamaguchi: "Image quality evaluation and control of computer-generated hologram with structural similarity index", The 2017 International Workshop on Advanced Imaging Technology, papar 1C–1 (Jan. 2017).
15) S. A. Benton, V. M. Bove Jr: Holographic Imaging, John Wiley & Sons, Inc. (2008).
16) M. E. Lucente: "Interactive computation of holograms using a look-up table", Journal of Electronic Imaging, Vol. 2, No. 1, pp. 28–34 (Jan. 1993).
17) L. B. Lesem, P. M. Hirsch, J. A. Jordan: "The Kinoform: A New Wavefront Recon struction Device", IBM Journal of Research and Development, Vol. 13, No. 2, pp. 150–155 (Mar. 1969).
18) R. W. Gerchberg, W. O. Saxton: "A practical algorithm for the determination of the phase from image and diffraction plane pictures", Optik, Vol. 35, No. 2, pp. 237–246 (Jan. 1972).
19) G. T. Nehmetallah, R. Aylo, L. Willams: "Analog and Digital Holography with MATLAB", SPIE Press (2015).
20) University of Southern California: "The USC-SIPI Image Database", http://sipi.usc.edu/database/

(Received October 24, 2023)
(Revised November 15, 2023)

Hiroshi YOSHIKAWA (*Member*)
He received his PhD in electrical engineering from Nihon University in 1985 and joined the faculty at Nihon University where he currently holds the position of Professor of Computer Engineering. Current researches are in computer-generated holograms, holographic printer, holographic video display and computer graphics.

Takeshi YAMAGUCHI
He received his PhD in electrical engineering from Nihon University. He is currently associate professor at Nihon University. His research interests include electroholography, computer-generated holograms, and computer graphics.

Call for Papers

Special Issue on

Image Electronics Technologies Related to VR/AR/MR/XR

IIEEJ Editorial Committee

In recent years, Virtual Reality (VR), Augmented Reality (AR), Mixed Reality (MR), Extended/Cross Reality (XR) have been applied to systems of various fields, and have brought revolutionary innovation in them. The applied area includes Entertainment, Medical & Health Care, Education, Practice & Disciplining, Manufacturing, Remote Work & Meeting, and so on, and has been extended even for the purposes of Art and Environment Preservation. Especially, VR technology will provide fully virtual environment and invite the users to another world in order to experience unfeasible experiments. Also, AR technology will improve the efficiency of daily task by shortening the access time to the desired information, and will empower the process to understand the individual utilized context at the same time.

On the other hand, the long-time effect of such technologies on human senses and mental states has been gathering wide interest, and how to coexist with such technologies will be the future big theme for human beings.

In this special issue, we invite various categories (Ordinary paper, Short paper, System development paper, Practice Oriented Paper) of papers on designing and implementing technologies that support VR/AR/MR/XR, and on the investigation and evaluation of such technologies and systems being developed. We look forward to receiving your contributions.

1. Topics covered include but not limited to
- Technologies and Systems on VR/AR/MR/XR, Meta-verse
- User Interface, Human Computer Interaction, Mutual Interaction
- Computer Graphics, Image Processing, Image Communication
- Computer Vision, Image Understanding, Pattern Recognition, Machine Learning, AI
- Collaboration, Education, Training, Medical Application, Remote Work/Meeting/Entertainment

2. Treatment of papers

Submission paper style format and double-blind peer review process are the same as the regular paper. If the number of accepted papers is less than the minimum number for the special issue, the acceptance paper will be published as the regular contributed paper. We ask for your understanding and cooperation.

3. Publication of Special Issue:

IIEEJ Transactions on Image Electronics and Visual Computing Vol.13, No.1 (June 2025)

4. Submission Deadline:

Tuesday, October 31, 2024

5. Contact details for Inquiries:

IIEEJ Office E-mail: hensyu@iieej.org

6. Online Submission URL: http://www.editorialmanager.com/iieej/

Revised: January 6, 2017

Revised: July 6, 2018

Guidance for Paper Submission

1. Submission of Papers

(1) Preparation before submission

- The authors should download "Guidance for Paper Submission" and "Style Format" from the "Academic Journals", "English Journals" section of the Society website and prepare the paper for submission.
- Two versions of "Style Format" are available, TeX and MS Word. To reduce publishing costs and effort, use of TeX version is recommended.
- There are four categories of manuscripts as follows:
  - Ordinary paper: It should be a scholarly thesis on a unique study, development or investigation concerning image electronics engineering. This is an ordinary paper to propose new ideas and will be evaluated for novelty, utility, reliability and comprehensibility. As a general rule, the authors are requested to summarize a paper within eight pages.
  - Short paper: It is not yet a completed full paper, but instead a quick report of the partial result obtained at the preliminary stage as well as the knowledge obtained from the said result. As a general rule, the authors are requested to summarize a paper within four pages.
  - System development paper: It is a paper that is a combination of existing technology or it has its own novelty in addition to the novelty and utility of an ordinary paper, and the development results are superior to conventional methods or can be applied to other systems and demonstrates new knowledge. As a general rule, the authors are requested to summarize a paper within eight pages.
  - Data Paper: A summary of data obtained in the process of a survey, product development, test, application, and so on, which are the beneficial information for readers even though its novelty is not high. As a general rule, the authors are requested to summarize a paper within eight pages.
- To submit the manuscript for ordinsry paper, short paper, system development paper, or data paper, at least one of the authors must be a member or a student member of the society.
- We prohibit the duplicate submission of a paper. If a full paper, short paper, system development paper, or data paper with the same content has been published or submitted to other open publishing forums by the same author, or at least one of the co-authors, it shall not be accepted as a rule. Open publishing forum implies internal or external books, magazines, bulletins and newsletters from government offices, schools, company organizations, etc. This regulation does not apply to a preliminary draft to be used at an annual meeting, seminar, symposium, conference, and lecture meeting of our society or other societies (including overseas societies). A paper that was once approved as a short paper and being submitted again as the full paper after completion is not regarded as a duplicate submission.

(2) Submission stage of a paper

- Delete all author information at the time of submission. However, deletion of reference information is the author's discretion.
- At first, please register your name on the paper submission page of the following URL, and then log in again and fill in the necessary information. Use the "Style Format" to upload your manuscript. An applicant should use PDF format (converted from dvi of TeX or MS Word

format) for the manuscript. As a rule, charts (figures and tables) shall be inserted into the manuscript to use the "Style Format". (a different type of data file, such as audio and video, can be uploaded at the same time for reference.)

http://www.editorialmanager.com/iieej/

- If you have any questions regarding the submission, please consult the editor at our office.

Contact:
Person in charge of editing
The Institute of Image Electronics Engineers of Japan
3-35-4-101, Arakawa, Arakawa-Ku, Tokyo 116-0002, Japan
E-mail: hensyu@iieej.org
Tel: +81-3-5615-2893, Fax: +81-3-5615-2894

2. Review of Papers and Procedures

(1) Review of a paper

- A manuscript is reviewed by professional reviewers of the relevant field. The reviewer will deem the paper "acceptance", "conditionally acceptance" or "returned". The applicant is notified of the result of the review by E-mail.
- Evaluation method

  Ordinary papers are usually evaluated on the following criteria:
  - ✓ Novelty: The contents of the paper are novel.
  - ✓ Utility: The contents are useful for academic and industrial development.
  - ✓ Reliability: The contents are considered trustworthy by the reviewer.
  - ✓ Comprehensibility: The contents of the paper are clearly described and understood by the reviewer without misunderstanding.

  Apart from the novelty and utility of an ordinary paper, a short paper can be evaluated by having a quickness on the research content and evaluated to have new knowledge with results even if that is partial or for specific use.

  System development papers are evaluated based on the following criteria, apart from the novelty and utility of an ordinary paper.
  - ✓ Novelty of system development: Even when integrated with existing technologies, the novelty of the combination, novelty of the system, novelty of knowledge obtained from the developed system, etc. are recognized as the novelty of the system.
  - ✓ Utility of system development: It is comprehensively or partially superior compared to similar systems. Demonstrates a pioneering new application concept as a system. The combination has appropriate optimality for practical use. Demonstrates performance limitations and examples of performance of the system when put to practical use.

  Apart from the novelty and utility of an ordinary paper, a data paper is considered novel if new deliverables of test, application and manufacturing, the introduction of new technology and proposals in the worksite have any priority, even though they are not necessarily original. Also, if the new deliverables are superior compared to the existing technology and are useful for academic and industrial development, they should be evaluated.

(2) Procedure after a review

- In case of acceptance, the author prepares a final manuscript (as mentioned in 3.).
- In the case of acceptance with comments by the reviewer, the author may revise the paper in consideration of the reviewer's opinion and proceed to prepare the final manuscript (as

mentioned in 3.).

- In case of conditional acceptance, the author shall modify a paper based on the reviewer's requirements by a specified date (within 60 days), and submit the modified paper for approval. The corrected parts must be colored or underlined. A reply letter must be attached that carefully explains the corrections, assertions and future issues, etc., for all of the acceptance conditions.
- In case a paper is returned, the author cannot proceed to the next step. Please look at the reasons the reviewer lists for the return. We expect an applicant to try again after reviewing the content of the paper.

(3) Review request for a revised manuscript

- If you want to submit your paper after conditional acceptance, please submit the reply letter to the comments of the reviewers, and the revised manuscript with revision history to the submission site. Please note the designated date for submission. Revised manuscripts delayed more than the designated date be treated as new applications.
- In principle, a revised manuscript willl be reviewed by the same reviewer. It is judged either accceptance or returned.
- After the judgment, please follow the same procedure as (2).

3. Submission of final manuscript for publication

(1) Submission of a final manuscript

- An author, who has received the notice of "Acceptance", will receive an email regarding the creation of the final manuscript. The author shall prepare a complete set of the final manuscript (electronic data) following the instructions given and send it to the office by the designated date.
- The final manuscript shall contain a source file (TeX edition or MS Word version) and a PDF file, eps files for all drawings (including bmp, jpg, png), an eps file for author's photograph (eps or jpg file of more than 300 dpi with length and breadth ratio 3:2, upper part of the body) for authors' introduction. Please submit these in a compressed format, such as a zip file.
- In the final manuscript, write the name of the authors, name of an organizations, introduction of authors, and if necessary, an appreciation acknowledgment. (cancel macros in the Style file)
- An author whose paper is accepted shall pay a page charge before publishing. It is the author's decision to purchase offprints. (ref. page charge and offprint price information)

(2) Galley print proof

- The author is requested to check the galley (hard copy) a couple of weeks before the paper is published in the journal. Please check the galley by the designated date (within one week). After making any corrections, scan the data and prepare a PDF file, and send it to our office by email. At that time, fill in the Offprint Purchase Slip and Copyright Form and return the scanned data to our office in PDF file form.
- In principle, the copyrights of all articles published in our journal, including electronic form, belong to our society.
- You can download the Offprint Purchase Slip and the Copyright Form from the journal on our homepage. (ref. Attachment 2: Offprint Purchase Slip, Attachment 3: Copyright Form)

(3) Publication

- After final proofreading, a paper is published in the Academic journal or English transaction (both in electronic format) and will also be posted on our homepage.

IIEEJ Trans. on Image Electronics and Visual Computing

Vol. 11, No. 2 2023

Editor in Chief: Osamu Uchida
The Institute of Image Electronics Engineers of Japan
3-35-4-101, Arakawa, Arakawa-ku, Tokyo 116-0002, Japan

Print: ISSN 2188-1898
Online: ISSN 2188-1901
CD-ROM: ISSN 2188-191x