1. Record Nr. UNINA9910847083303321

Autore	Zhang Fang-Lue
Titolo	Computational Visual Media : 12th International Conference, CVM 2024, Wellington, New Zealand, April 10–12, 2024, Proceedings, Part I / / edited by Fang-Lue Zhang, Andrei Sharf
Pubbl/distr/stampa	Singapore : , : Springer Nature Singapore : , : Imprint : Springer, , 2024
ISBN	981-9720-95-8
Edizione	[1st ed. 2024.]
Descrizione fisica	1 online resource (331 pages)
Collana	Lecture Notes in Computer Science, , 1611-3349 ; ; 14592
Altri autori (Persone)	SharfAndrei
Disciplina	006.37
Soggetti	Computer vision Pattern recognition systems Application software Computer graphics Artificial intelligence Algorithms Computer Vision Automated Pattern Recognition Computer and Information Systems Applications Computer Graphics Artificial Intelligence
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Organization -- Contents - Part I -- Contents - Part II -- Reconstruction and Modelling -- PIFu for the Real World: A Self-supervised Framework to Reconstruct Dressed Human from Single-View Images -- 1 Introduction -- 2 Related Work -- 2.1 Singe-View Human Reconstruction -- 2.2 Single-View Depth Estimation -- 2.3 Self-supervised 3D Reconstruction -- 3 Method -- 3.1 Normal and Depth Estimation -- 3.2 SDF-Based Pixel-Aligned Implicit Function from Depth -- 3.3 Depth-Guided Self-supervised Learning -- 4 Experiments -- 4.1 Datasets, Metrics, and Implementation Details -- 4.2 Evaluations -- 4.3 Comparison with the State-of-the-Art -- 5 Conclusion -- References -- Sketchformer++: A Hierarchical

Transformer Architecture for Vector Sketch Representation -- 1 Introduction -- 2 Related Work -- 3 Method -- 3.1 Data Representation -- 3.2 Hierarchical Transformer Architecture -- 3.3 Training -- 4 Experiments -- 4.1 Sketch Reconstruction -- 4.2 Sketch Recognition -- 4.3 Sketch Semantic Segmentation -- 4.4 Ablation Study -- 5 Conclusion -- References -- Leveraging Panoptic Prior for 3D Zero-Shot Semantic Understanding Within Language Embedded Radiance Fields -- 1 Introduction -- 2 Related Works -- 2.1 NeRF with Semantics -- 2.2 Panoptic Segmentation -- 2.3 Open-Vocabulary Object Detection -- 2.4 Zero-Shot Learning in 3D -- 2.5 Cross-Modal Knowledge Distillation -- 3 Method -- 3.1 Overview -- 3.2 Field Structure -- 3.3 Semantic Prior Extraction -- 3.4 CLIP Pyramid Reconstruction -- 3.5 Relevancy Evaluation Metric -- 4 Experiments -- 4.1 Settings -- 4.2 Qualitative Results -- 4.3 Ablation Study -- 5 Limitations -- 6 Conclusions -- References -- Multi-Scale Implicit Surface Reconstruction for Outdoor Scenes -- 1 Introduction -- 2 Related Work -- 3 Method -- 3.1 Multi-scale Rendering with SDF Representation -- 3.2 Dynamic Position Encoding.
3.3 Adaptive Sampling Strategy in Image Space -- 3.4 More Details -- 3.5 Loss Function -- 4 Experiments -- 4.1 Implementation Details -- 4.2 Qualitative and Quantitative Comparisons -- 4.3 Ablation Study -- 5 Conclusion -- References -- Neural Radiance Fields for Dynamic View Synthesis Using Local Temporal Priors -- 1 Introduction -- 2 Related Work -- 3 Overview -- 4 Dynamic Scene Representation -- 5 Local Temporal NeRF -- 5.1 Local Temporal Module -- 5.2 Loss Functions -- 5.3 Implementation -- 6 Results -- 6.1 Quantitative Evaluation -- 6.2 Qualitative Evaluation -- 6.3 Ablation Study -- 6.4 Additional Comparisons -- 7 Limitations and Discussion -- 8 Conclusion -- References -- Point Cloud -- Point Cloud Segmentation with Guided Sampling and Continuous Interpolation -- 1 Introduction -- 2 Related Work -- 2.1 Point Cloud Learning -- 2.2 Point Cloud Sampling -- 3 Method -- 3.1 Motivation -- 3.2 Guided Sampling -- 3.3 Continuous Interpolation -- 4 Experiments -- 4.1 Experimental Setup -- 4.2 Signal Reconstruction -- 4.3 Semantic Segmentation -- 4.4 Object Part Segmentation -- 4.5 Ablation Study -- 5 Conclusion and Discussion -- References -- TopFormer: Topology-Aware Transformer for Point Cloud Registration -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 Problem Definition -- 3.2 Local Feature Encoder -- 3.3 Topology-Aware Transformer -- 3.4 Sparse Point Matching -- 3.5 Dense Points Refinement -- 3.6 Loss Function -- 4 Experiments -- 4.1 Implementation -- 4.2 Indoor Scene: 3DMatch -- 4.3 Outdoor Scene Data: KITTI -- 4.4 Ablation Study -- 5 Conclusion -- References -- Adversarial Geometric Transformations of Point Clouds for Physical Attack -- 1 Introduction -- 2 Related Works -- 3 Methodology -- 3.1 Preliminaries -- 3.2 Adversarial Geometric Transformations -- 3.3 Optimization -- 4 Experiments -- 4.1 Dataset and Settings.
4.2 Evaluation on Adversarial Point Clouds -- 4.3 Evaluation on Shape and Physical Attack -- 4.4 Ablation Studies -- 5 Conclusions -- References -- SARNet: Semantic Augmented Registration of Large-Scale Urban Point Clouds -- 1 Introduction -- 2 Related Work -- 2.1 Traditional Feature-Based Registration -- 2.2 Learning-Based Registration -- 2.3 3D Point Feature Learning -- 3 Problem Statement and Overview -- 4 Methodology -- 4.1 Semantic-Based Farthest Point Sampling -- 4.2 Semantic-Augmented Feature Extraction -- 4.3 Semantic-Refined Transformation Estimation -- 4.4 Loss Functions -- 4.5 Implementation Details -- 5 Experimental Results -- 5.1 Experimental Setup -- 5.2 Evaluation Metrics -- 5.3 Comparisons -- 5.4 Ablation Study -- 5.5 Limitations -- 6 Conclusion and Future Work

-- References -- Rendering and Animation -- FASSET: Frame Supersampling and Extrapolation Using Implicit Neural Representations of Rendering Contents -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 Motivation and Overview -- 3.2 Implicit Neural Representations of Rendering Contents -- 3.3 Frame Feature Extractor -- 3.4 Network Training -- 4 Experiments -- 4.1 Dataset -- 4.2 Baselines and Settings -- 4.3 Analysis of Runtime Performance and Model Efficiency -- 4.4 Ablation Study -- 4.5 Limitation -- 5 Conclusion -- References -- MatTrans: Material Reflectance Property Estimation of Complex Objects with Transformer -- 1 Introduction -- 2 Related Works -- 3 Method -- 3.1 Initial Estimation Network -- 3.2 Refined Estimation Network -- 3.3 Transformer Encoder -- 3.4 Dataset -- 3.5 Training -- 4 Experiments -- 4.1 Ablation Experiment -- 4.2 Generalization to Real Data -- 4.3 Comparison Experiment -- 5 Conclusion -- References -- Improved Text-Driven Human Motion Generation via Out-of-Distribution Detection and Rectification -- 1 Introduction -- 2 Related Work.
3 The Proposed Method -- 4 Experiments -- 4.1 Dataset and Evaluation Metrics -- 4.2 Experiment Configuration and Training Details -- 4.3 Comparisons Between Different Text-Driven Human Motion Generation Methods -- 4.4 Comparison Between Different Outlier Detection Algorithms -- 4.5 Evaluation of Different Thresholds for Outlier Detection -- 4.6 Ablation Study -- 5 Conclusion -- References -- User Interactions -- BK-Editer: Body-Keeping Text-Conditioned Real Image Editing -- 1 Introduction -- 2 Related Work -- 3 Background -- 3.1 Diffusion Model Training -- 3.2 DDIM Sampling and Inversion -- 3.3 Text Condition and Classifier-Free Guidance -- 3.4 Stable Diffusion Model -- 3.5 Task Setting and the Body-Keeping Problem -- 4 Method -- 4.1 Tuning Stage for Finetuning Network -- 4.2 Inversion Stage for Obtaining BK-Attn Embeddings -- 4.3 Edit Stage with Body-Keeping -- 5 Experiments -- 5.1 Comparisons with Other Concurrent Works -- 5.2 User Study -- 5.3 Ablation Study -- 6 Limitations and Conclusion -- References -- Walking Telescope: Exploring the Zooming Effect in Expanding Detection Threshold Range for Translation Gain -- 1 Introduction -- 2 Related Work -- 2.1 Translation Gain Detection Threshold -- 2.2 Impact of FoV Change on Distance Perception -- 2.3 Impact of Magnified View on Distance Perception -- 3 Method -- 3.1 Translation Gain -- 3.2 Motivation -- 3.3 Verification Experiment -- 4 Main Experiment -- 4.1 Design and Hypotheses -- 4.2 Apparatus -- 4.3 Participants -- 4.4 Procedure -- 5 Results -- 5.1 Direction Thresholds -- 5.2 Simulator Sickness -- 6 Discussion -- 7 Limitation and Future Work -- 8 Conclusion -- References -- A U-Shaped Spatio-Temporal Transformer as Solver for Motion Capture -- 1 Introduction -- 2 Related Work -- 2.1 MoCap Data Clean-Up and Solving -- 2.2 Smoothness -- 2.3 Rotation Representations -- 2.4 Attention Model.
2.5 U-Net Architecture -- 3 Methodology -- 3.1 Problem Formulation -- 3.2 Overall Structure -- 4 Experiments and Evaluation -- 4.1 Experimental Settings -- 4.2 Quantitative and Qualitative Research -- 4.3 Ablation Study -- 5 Limitations and Future Work -- 6 Conclusion -- References -- ROSA-Net: Rotation-Robust Structure-Aware Network for Fine-Grained 3D Shape Retrieval -- 1 Introduction -- 2 Related Work -- 2.1 3D Shape Retrieval -- 2.2 Mesh-Based Representations -- 2.3 Rotation-Invariant Representations -- 3 ROSA-Net -- 3.1 Overview -- 3.2 Geometric Feature Representation -- 3.3 Part Geometry Attention Mechanism -- 3.4 Structural Information Representation -- 3.5 Geometry-Structure Attention Mechanism -- 3.6 Global Feature Encoding -- 3.7 Losses -- 3.8 Model Training and Shape Retrieval -- 4

	Experimental Results -- 4.1 ROSA-Dataset -- 4.2 Fine-Grained Shape Retrieval -- 4.3 Weighted Features of Parts by Part-Geo Attention -- 4.4 Weighted Features by Geo-Struct Attention -- 4.5 Using Other Data Representation -- 4.6 Ablation Study -- 5 Conclusion -- References -- Author Index.
Sommario/riassunto	This book constitutes the refereed proceedings of CVM 2024, the 12th International Conference on Computational Visual Media, held in Wellington, New Zealand, in April 2024. The 34 full papers were carefully reviewed and selected from 212 submissions. The papers are organized in topical sections as follows: Part I: Reconstruction and Modelling, Point Cloud, Rendering and Animation, User Interations. Part II: Facial Images, Image Generation and Enhancement, Image Understanding, Stylization, Vision Meets Graphics.