Vision-Based Multimodal Interfaces: Classification and Insights for Context-Aware System Design

Macro-Level

Context of VMIs

Context Source Factors

System Factors · Human Factors · Environmental Factors

Context Categories

Activity · Location · Identity · Time

Micro-Level

Input Data Modality
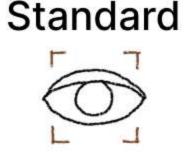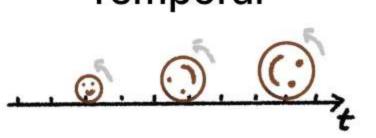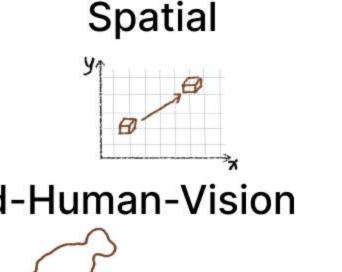
Visual Modality and Visual Dimensions

Standard · Scale · Spatial

Temporal · Beyond-Human-Vision

Other Modalities

Audio · Haptic · Positional

Text · Physiological · Motion

System Design Foundations
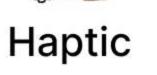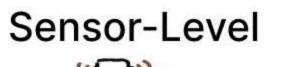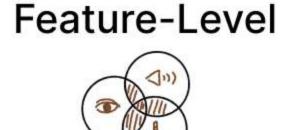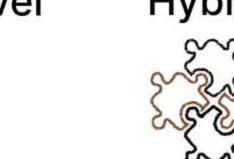
Data Integration Stages

Sensor-Level · Feature-Level

Information-Level · Hybrid

Multimodal Data Processing

APIs · ML Models · Heuristic
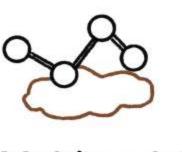
Evaluation Strategies

Demo · Technical · User

Macro-Level

Key Challenges

Privacy & Security · User Variability · Ethics · Cognitive Load & Engagement · Sensor Automation · Context Annotation
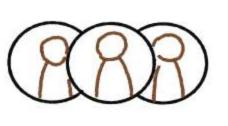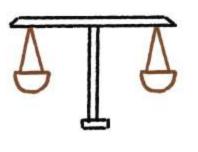
Multimodal Integration · Contextual Reasoning · Data Imbalance · Assessment Heterogeneity · Scalable Architecture · Power Efficiency
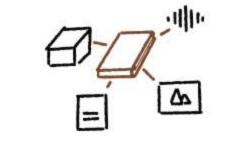
Application Domains

Local & Identity Recognition · Accessibility · Autonomous and Assistive Driving · Education · Game · Content Retrieval, Editing & Creation
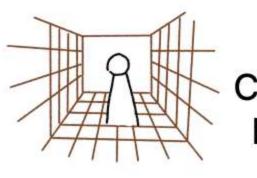
Spatial Computing & Perception · Well-being & Health Care · Activity Detection & Understanding