RASTER DOMAIN TEXT STEGANOGRAPHY: A UNIFIED GLYPH LEVEL FRAMEWORK FOR MULTIMODAL SECURE EMBEDDING

UDAY KIRAN KANDALA, MSC STUDENT, QUEEN MARY UNIVERSITY OF LONDON

ABSTRACT:

This work introduces a raster-domain text steganography method, Glyph Perturbation Cardinality (GPC), for embedding heterogeneous data directly within rendered text images. Unlike linguistic or formatting-based text steganography, the proposed approach operates entirely in the pixel domain and treats each glyph as a bounded carrier unit. A payload value is encoded as the number of minimally perturbed interior ink pixels within a glyph, producing visually imperceptible modifications while preserving overall appearance. Decoding is performed through direct differential analysis between the cover image and the stego image, where pixel-wise differences within glyph regions reveal the perturbation cardinality and recover the embedded data. The framework supports lossless text-to-text embedding and extends to multimodal payloads by mapping image, audio, and video signals into bounded integer sequences distributed across glyphs. The results demonstrate that rasterized text images can function as a lightweight and visually covert carrier independent of linguistic structure, layout manipulation, or text encoding.

INTRODUCTION:

Text steganography has historically embedded information by manipulating the text itself through whitespace and layout encoding, synonym substitution, lexical constraints, syntactic transformations, paragraph resizing, splitting and merging, Unicode and homoglyph substitution, font shaping, location-index (zero-distortion) schemes, OCR-error induction, probabilistic language models such as Markov chains, LSTM generation, large language model based generation with semantic or emotional control, entropy-constrained sampling, black-box keyword mapping, hybrid linguistic-visual encoding, and cross-modal semantic pairing [1]. Although diverse, these approaches remain tied to the logical structure of language and are therefore sensitive to rewriting, normalization, translation, or automated processing, while also limiting embedding capacity.

In this work we step away from language entirely and operate after text rendering. Once rasterized, each character becomes a small pixel container rather than a linguistic symbol. We encode information by slightly modifying a controlled number of interior ink pixels inside each glyph. These changes are visually imperceptible but recoverable by comparing localized pixel differences under the same rendering conditions. Because the method works on the rendered bitmap rather than words, grammar, or encoding, it remains independent of vocabulary and formatting and can carry different types of payloads such as text, images, audio, and video within ordinary documents.
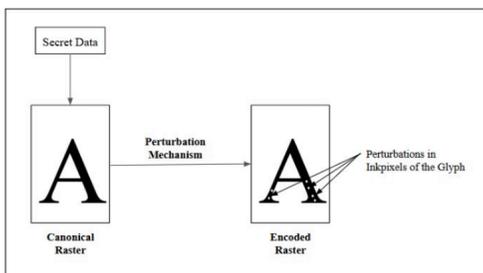

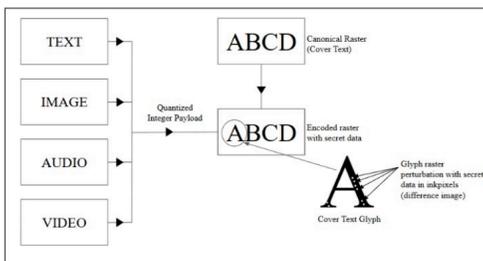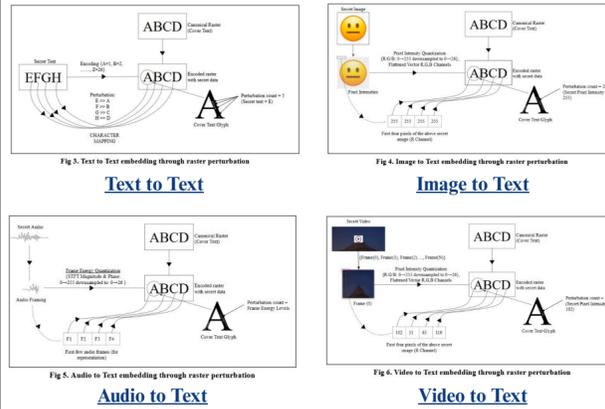Fig 1. Perturbating secret data in canonical raster

Fig 2. Multimodal data embedding into text raster glyphs

MECHANISM:

Hetrogenous data is embedded into the Glyph Rasters of Texts using the proposed Glyph Perturbation Cardinality Mechanism.

Fig 3. Text to Text embedding through raster perturbation

Text to Text

Fig 4. Image to Text embedding through raster perturbation

Image to Text

Fig 5. Audio to Text embedding through raster perturbation

Audio to Text

Fig 6. Video to Text embedding through raster perturbation
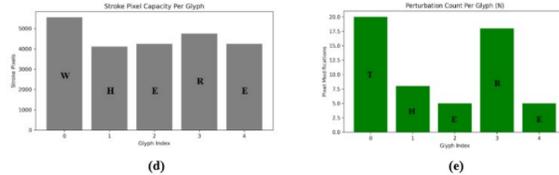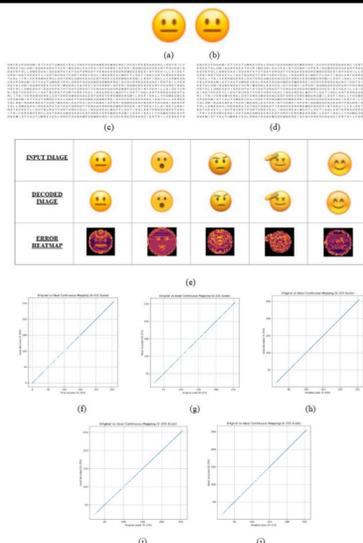
Video to Text

Text to Text:

Image to Text:

Using the proposed method, five different emoji images are quantized and pertubed into the textual rasters and decoded accordingly. The metrics such as MAE, MSE, RMSE, PSNR, SSIM and Heatmaps, shows a promising encoding and decoding of the images into textual rasters.

METRICS	IMAGE 1	IMAGE 2	IMAGE 3	IMAGE 4	IMAGE 5
MSE	4.2161	5.3375	5.2438	4.1809	4.8157
MAE	1.2916	1.7014	1.5856	1.2526	1.4231
RMSE	2.0533	2.3103	2.2899	2.0447	2.1944
PSNR	41.8816	40.857	40.9343	41.918	41.3041
SSIM AVG	0.986	0.989	0.9928	0.9929	0.9914
SSIM R	0.9666	0.9869	0.9907	0.988	0.987
SSIM G	0.997	0.9936	0.9971	0.9977	0.9957
SSIM B	0.993	0.9876	0.9906	0.9931	0.9913

Table1. Metrics for Image to Text raster perturbations.

Audio to Text:

Five different audio samples are used in the experimentation, where each audio file is sampled and quantized to 26 levels and are embedded into the textual glyphs. The decoded values are reverse mapped into the original signals to acquire the input audio file. The metrics are promising but show some noise due to coarse quantization technique used during experiments.
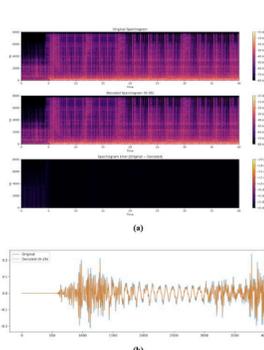
The metrics shows better MAE and MSE values, though the Signal to Noise ratio of the decoded audio is in acceptable range of 24-28, better quantization or pre and post processing can result in better decoded audio re-construction.

Video to Text:

Column 1	MAE	MSE	SNR
Audio1	0.00721	0.00011224	28.288
Audio2	0.00915	0.000185	24.237
Audio3	0.00432	0.00003146	28.058
Audio4	0.006	0.00155	26.489
Audio5	0.00537	0.00004796	27.286

Table 2. Metrics for Audio to Text raster perturbations.

Different videos are used in the experiements to test the proposed mechanism, the videos are of different textures, such as mountains, oceans, animals, people and abstract paintings.
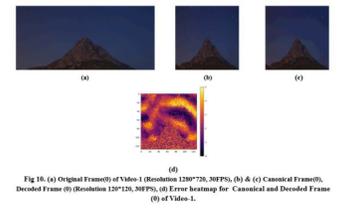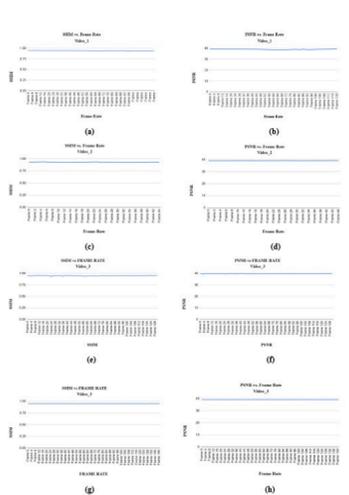
The resolution of all the videos are reduced to 128*128 and are embedded into the Glyph rasters and the decoding shows accurate retrieval of the structural information and motion of the content without any deviations or noise. The metrics shows a promising SSIM and PSNR values of the decoded frames. Better quanitzation techniques can further improve the quality of the videos decoded and can result in better data transmission.
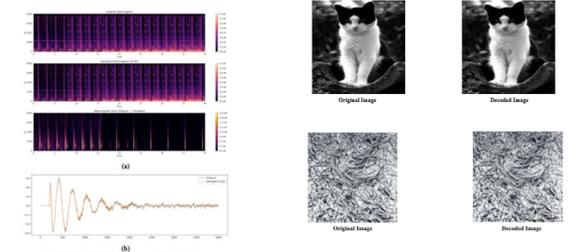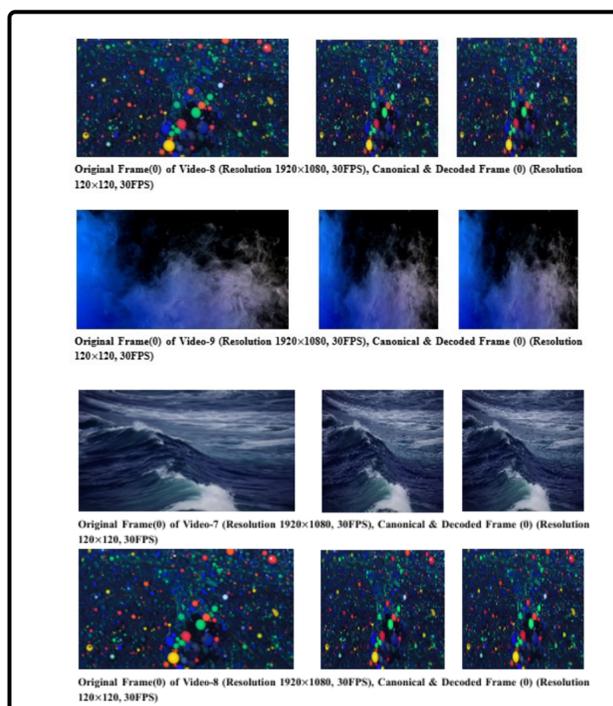

Fig 10. (a) Original Frame(0) of Video 1 (Resolution 128*720, 30FPS), (b) & (c) Canonical Frame(0), Decoded Frame (0) (Resolution 120*128, 30FPS), (d) Error heatmap for Canonical and Decoded Frame (0) of Video 1.

Additional Results:

Fig 11: Audio-10. (a) Spectrograms and (b) Waveform

Original Frame(0) of Video-8 (Resolution 1920×1080, 30FPS), Canonical & Decoded Frame (0) (Resolution 120×120, 30FPS)

Original Frame(0) of Video-9 (Resolution 1920×1080, 30FPS), Canonical & Decoded Frame (0) (Resolution 120×130, 30FPS)

Original Frame(0) of Video-7 (Resolution 1920×1080, 30FPS), Canonical & Decoded Frame (0) (Resolution 120×120, 30FPS)

Original Frame(0) of Video-8 (Resolution 1920×1080, 30FPS), Canonical & Decoded Frame (0) (Resolution 120×120, 30FPS)

CONCLUSION:

This work demonstrates that rendered text can act as a universal carrier for hidden information when treated as a geometric pixel field rather than a linguistic object. By encoding data through the number of minimally perturbed interior glyph pixels, the proposed perturbation cardinality channel enables reliable recovery of text and preserves strong perceptual quality for images, video, and audio. The observed limitations stem mainly from the current 26-level quantization and can be expanded through richer encoding and modality-aware processing without changing the core raster-domain principle. Overall, the results establish deterministic glyph rasterization as a practical foundation for multimodal steganography, showing that ordinary text can transport structured data with high visual imperceptibility and predictable reconstruction.