

REIST Image

Regionally Enhanced Incremental Scalable Transmission
Ein wahrnehmungsorientiertes Verfahren fuer progressive Bilduebertragung

Rudolf Stepan (OE3SRO)
Forschungs- und Implementierungsbericht, Version 2

29. September 2026

Zusammenfassung

Dieses Papier beschreibt REIST IMAGE, *Regionally Enhanced Incremental Scalable Transmission*, ein experimentelles Verfahren zur progressiven Bilduebertragung unter begrenzter oder schwankender Uebertragungskapazitaet. Gegenueber dem urspruenglichen Konzept liegt nun eine Referenzimplementierung vor. Sie sendet zunaechst Layout und Regionsindex, danach eine entropiecodierte Vektorskizze, einen kleinen Raster-Basislayer, globale Verfeinerungen und schliesslich dynamisch priorisierte regionale Verfeinerungen. Ein Referenzlauf auf einer synthetischen Szene mit 640×400 Pixeln erreichte SSIM 0,808 nach 1.102 Byte und SSIM 0,954 nach 4.004 Byte, entsprechend rund 2,0 % des 196.126-Byte-Streams. Diese Werte sind vorlaeufige Implementierungsmessungen, kein Benchmark: Der damalige Mess-Harness ist nicht im Repository enthalten und die Stichprobe ist klein. Das Papier dokumentiert Architektur, Vektorskizze, Scheduler, erste Validierung sowie die notwendigen Schritte fuer eine reproduzierbare Evaluation. **Schlüsselwoerter:** progressive Bilduebertragung, scalable coding, perceptual quality, adaptive scheduling, Web, SDR, low bandwidth

1 Motivation

Bei einer klassischen Dateiuebertragung ist der logische Erfolg haeufig binaer: vollstaendig oder unvollstaendig. Fuer Bilder ist dies bei langsamen Kanaelen unguenstig. Progressive JPEG kennt progressive und hierarchische Betriebsarten [1]; JPEG 2000 wurde fuer skalierbare und interaktive Bildrepraesentation entwickelt [2, 3]. REIST IMAGE setzt deshalb nicht bei der blossen Idee “unscharf zuerst, scharf spaeter” an. Forschungsgegenstand

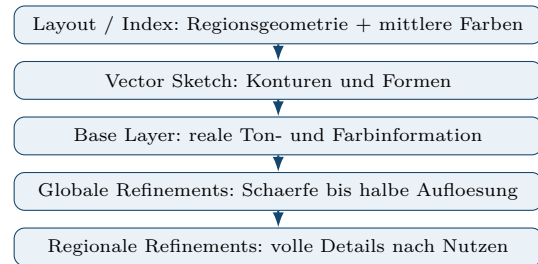


Abbildung 1: Aktuelle fuefnstufige REIST-Image-Pipeline.

ist die *Priorisierung inkrementeller visueller Information*.

2 Begriff und Abgrenzung

REIST IMAGE steht fuer **Regionally Enhanced Incremental Scalable Transmission**. Der Name bezeichnet eine Format- und Uebertragungsfamilie: Regionen koennen unterschiedlich priorisiert werden; neue Daten verbessern einen bereits sichtbaren Zustand; Qualitaet und Aufloesung skalieren; Reihenfolge und Scheduling sind Teil des Verfahrens. Der Entwurf ist bewusst nicht an Residualcodierung, Wavelets oder eine einzelne Transformation gebunden.

3 Aktueller Implementierungsstand

Die Referenzimplementierung zerlegt das Zielbild content-adaptiv mit einem Quadtree. Regionen werden dort weiter geteilt, wo relative Kantenenergie oder Kontrast gegenueber den globalen Bildstatistiken hoch sind. Die aktuelle Implementierung

begrenzt die Endbreite auf 640 Pixel, erzwingt mindestens ein 4×4 -Grundraster und verwendet eine minimale Regionsgrosse von 20 Pixeln. Dadurch entstehen bei typischen Fotografien wenige grosse Regionen in ruhigen Flaechen und viele kleine Regionen ueber Gesichtern, Text und Kanten.

Die ersten vier Stufen bilden einen festen Prefix; erst die regionalen Bloecke werden dynamisch geplant:

Layout $\rightarrow$ Sketch $\rightarrow$ Base $\rightarrow$ Global $\rightarrow$ Regional.

Ein Abbruch hinter jedem vollstaendig empfangenen Block hinterlaesst einen darstellbaren Zwischenzustand.

3.1 Layout / Index

Der erste Block uebertraegt pro Region Position, Ausdehnung und mittlere RGB-Farbe. Der Decoder kann damit sofort eine vollflaechige abstrakte Komposition rendern. Im dokumentierten Referenzlauf kostete dieser Block 1.102 Byte und erreichte bereits SSIM 0,808.

3.2 Vector-Sketch-Layer

Vor den Rasterverfeinerungen wird eine SVG-artige Konturskizze uebertragen. Die Referenzimplementierung bildet Luminanz, glaettet das Bild, bestimmt Sobel-Gradienten, wendet Non-Maximum-Suppression und Hysteresis an, verfolgt Konturen und vereinfacht sie mit Douglas–Peucker. Die resultierenden Polylinien werden mit einem adaptiven binaren Range-Coder gespeichert.

Der Sketch soll semantisch aussagekraeftige Geometrie frueh sichtbar machen. In den bisherigen Messungen veraenderte er den SSIM-Wert praktisch nicht, obwohl reale Testfotos bereits aus weniger als 4 KB Vektordaten visuell identifizierbare Personen, Fahrzeuge und Gebaeudestrukturen zeigten. Dies motiviert eine spaetere *Time-to-Recognition*-Metrik zusaetzlich zu PSNR und SSIM.

4 Formales Zielmodell

Sei I^* das Zielbild und $\hat{I}(t)$ der beim Empfaenger rekonstruierte Zustand. Der fruehe Zustand

Tabelle 1: Abstrakte Metadaten eines Refinement-Blocks.

Feld	Bedeutung
Level	Refinement-Stufe
Region	Position und Ausdehnung
Parent	benoetigte Vorgaengerinformation
Codec	Codierung des Payloads
Cost	Bytekosten
Priority	Nutzenhinweis
Integrity	CRC/Hash je Transportprofil

kann als $\hat{I}_{\text{early}} = \mathcal{R}(L, V, C)$ aufgefasst werden, wobei L Layout/Regionen, V Vektorpfade, C grobe Farbinformation und $\mathcal{R}$ den Renderer bezeichnet. $Q(\hat{I}, I^*)$ bezeichnet eine wahrnehmungsbezogene Qualitaetsfunktion. SSIM kann als strukturbezogene Metrik dienen [4]. Fuer die uebertragene Datenmenge $B(t)$ wird vereinfacht optimiert:

$$\max_{\pi} \frac{Q(\hat{I}_{\pi}(t), I^*) - Q(\hat{I}_0, I^*)}{B(t)}. \quad (1)$$

π ist die Reihenfolge der Refinement-Bloecke. Zusaetzlich soll die Qualitaet moeglichst monoton wachsen:

$$Q(\hat{I}_{k+1}, I^*) \geq Q(\hat{I}_k, I^*) - \varepsilon. \quad (2)$$

5 Datenmodell

Nach Layout und Sketch wird fuer die Rasterverfeinerung aus I^* eine Aufloesungspyramide $I_0, \dots, I_N$ gebildet, beispielsweise

$$32 \times 24 \rightarrow 64 \times 48 \rightarrow 128 \times 96 \rightarrow 256 \times 192 \rightarrow 512 \times 384.$$

I_0 ist der Base Layer. Eine erste Referenzimplementierung kann Residuen verwenden:

$$R_n = I_n - \text{Upscale}(I_{n-1}), \quad (3)$$

$$\hat{I}_n = \text{Upscale}(\hat{I}_{n-1}) + \hat{R}_n. \quad (4)$$

Die Residuen werden quantisiert, komprimiert und in unabhangig adressierbare Regionen zerlegt. Dies ist ein Startpunkt, keine normative Festlegung.

6 Regionale Priorisierung

Fuer jeden sendbaren Block b wird ein Score berechnet:

$$P(b) = \frac{\Delta Q(b)}{C(b)} W_c(b), \quad (5)$$

mit erwartetem Qualitaetsgewinn $\Delta Q(b)$, Bytekosten $C(b)$ und Kontextgewicht $W_c(b)$. Schaetzer koennen Kantenenergie, lokale Varianz, Text-/Gesichtserkennung oder eine gemessene Referenzmetrik kombinieren. Ein Alterungsfaktor verhindert das dauerhafte Verhungern ruhiger Regionen:

$$P'(b) = P(b)[1 + \lambda A(b)].$$

7 Webspezifisches Scheduling

Im Web kann $W_c = W_{\text{viewport}} W_{\text{interaction}} W_{\text{policy}}$ dynamisch angepasst werden. Sichtbare Bilder werden verfeinert, ausserhalb des Viewports kann der Base Layer genuegen; Zoom kann einzelne Regionen hochstufen. HTTP besitzt bereits standardisierte Prioritaetssignale. RFC 9218 definiert *urgency* und ein *incremental*-Flag fuer Antworten, deren Teile vor Abschluss Nutzen liefern [5]. REIST IMAGE kann solche Mechanismen nutzen, bleibt aber transportunabhaengig.

8 Referenzalgorithmus

Der Encoder normalisiert das Eingabebild, erzeugt die Pyramide, codiert den Base Layer, bestimmt Refinements, zerlegt sie regional, berechnet Kosten und Nutzen, speichert Abhaengigkeiten als gerichteten azyklischen Graphen und serialisiert Index und Payloads. Zur Laufzeit betrachtet der Scheduler nur Bloecke mit erfuellten Abhaengigkeiten und waehlt den hoechsten Score unter Beruecksichtigung von Fairness und aktuellem Kontext. Der Decoder rendert nach dem Base Layer sofort. Jeder gueltige Refinement-Block wird atomar angewendet; fehlende Bloecke duerfen unabhaengige Regionen nicht blockieren.

9 Container und Transport

Ein moeglicher Container besteht aus Header, frueh liegendem Index, Base Layer und Refinement-Bloeken. Der Index soll Range Requests oder

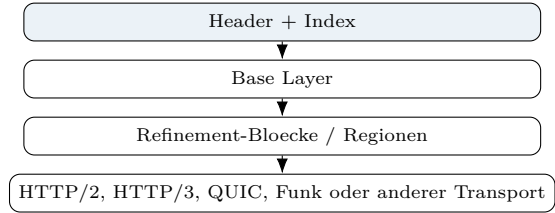


Abbildung 2: Formatlogik und Transport bleiben getrennte Schichten.

Tabelle 2: Vorlaeufige Prefix-Messung der Referenzimplementierung.

Stufe	kum. Byte	PSNR	SSIM
Layout / Index	1.102	22,7	0,808
Vector Sketch	1.399	22,7	0,808
Base Layer	2.368	27,1	0,855
Global L1	4.004	33,5	0,954
Global L2	7.519	41,4	0,991
Global L3	12.194	43,8	0,992

unabhaengige Streams erlauben. Die normative Byte-Syntax ist Gegenstand einer spaeteren Spezifikation.

10 Vorlaeufige Ergebnisse

Ein dokumentierter Referenzlauf auf einer synthetischen Szene mit 640×400 Pixeln erzeugte 100 Regionen, 247 Bloecke und insgesamt 196.126 Byte.

Damit wurden im Referenzlauf SSIM 0,80 nach 1.102 Byte (rund 0,6 % des Gesamtstreams) und SSIM 0,95 nach 4.004 Byte (rund 2,0 %) erreicht. Ueber alle 247 angewendeten Bloecke wurden in diesem Lauf keine Qualitaetsregressionen protokolliert.

Die Sketch-Geometrie wurde beim Serialisieren und Deserialisieren bitidentisch rekonstruiert; fuer den Range-Coder wurden 200 von 200 randomisierten Round-Trip-Tests verlustfrei dokumentiert. Fuer drei Sketch-Beispiele wurden 280, 2.707 und 3.607 Byte gemessen. Diese Ergebnisse belegen die Funktionsfaehigkeit des Prototyps, erlauben aber noch keine Aussage ueber eine Ueberlegenheit gegenueber Progressive JPEG oder JPEG 2000.

Tabelle 3: Vorgeschlagene experimentelle Stufen.

Phase	Abnahmekriterium
A	Base Layer sofort sichtbar; deterministische Refinements
B	Residuen vermeiden redundante Vollbilder
C	Adaptive Reihenfolge uebertrifft Zufallsreihenfolge bei gleichem Budget
D	Reproduzierbare Bandbreitensimulation mit Zeit/Byte/Qualitaets-Logs
E	Web-Prototyp mit inkrementellem Rendering und Viewport-Prioritaet
F	Vergleich gegen etablierte Verfahren ohne unbelegte Leistungsbehauptung
G	Stabilisierung von Container, Transportprofil und Datensatz

10.1 Reproduzierbarkeitsgrenze

Die Messwerte wurden mit einem temporaeren Konsolen-Harness erzeugt, der vor dem initialen Repository-Commit entfernt wurde. Sie sind deshalb aus einem sauberen Checkout derzeit nicht vollstaendig reproduzierbar. Ausserdem stammen sie aus einer kleinen Bildmenge und im Wesentlichen einer Zielaufloesung. Die Wiederherstellung eines versionierten Testprojekts und Benchmarks auf Standarddatensaetzen sind daher Voraussetzung fuer belastbare Leistungsbehauptungen.

11 Evaluationsplan

Der Vergleich darf nicht nur die Enddateigroesse betrachten. Zu messen sind Bytes und Zeit bis zur ersten sinnvollen Gesamtdarstellung, Zeit bis zu definierten Qualitaetsstufen, SSIM/PSNR, Decoderzeit, Speicherverbrauch, Qualitaetsgewinn pro KiB sowie subjektive *Time to Recognition*. Netzprofile sollen 64, 128 und 256 kbit/s sowie 1 und 5 Mbit/s, schwankende Bandbreite, Latenz und Paketverlust umfassen.

Verglichen werden mindestens: (1) sequentielle Uebertragung, (2) etablierte progressive Verfahren, (3) globale REIST-Refinements, (4) regionale Priorisierung und (5) Viewport-Priorisierung.

12 Funk- und Webprofil

Die Idee entstand aus dem Problem langsamer Funkbilduebertragung: Ein unvollstaendiges konventionelles Bild kann wertlos sein, waehrend ein frueher globaler Zustand bereits Information traegt. Im Funkprofil stehen robuste Pakete, FEC, Wiederholung und Unterbrechbarkeit im Vordergrund. Im Webprofil kommen Viewport, parallele Streams und Client-Feedback hinzu. Beide Profile teilen denselben abstrakten Refinement-Graphen.

13 Risiken und offene Fragen

Die groesste wissenschaftliche Gefahr waere, bekannte progressive Codierung lediglich neu zu benennen. Der eigenstaendige Beitrag muss deshalb experimentell zeigen, ob die *regionale, kontextabhaengige Schedulingstrategie* bei gleichem Bytebudget einen messbaren Vorteil fuer fruehe Nutzbarkeit liefert. Weitere Risiken sind Index-Overhead, Decoderkomplexitaet, sichtbare lokale Update-Artefakte, Fehlpriorisierung semantisch wichtiger Regionen sowie Cache- und CDN-Verhalten.

SSIM bildet strukturelle Aehnlichkeit besser ab als reine Fehlerenergie [4], ist aber nicht identisch mit “Motiv erkannt”. Deshalb ist eine kontrollierte Nutzerstudie fuer *Time to Recognition* vorgesehen.

14 Forschungsfragen

1. Senkt REIST IMAGE die Zeit bis zur sicheren Motiverkennung bei engem Bytebudget?
2. Welche Nutzenfunktion liefert den groessten Qualitaetsgewinn pro Byte?
3. Wie viel Nutzen bringt regionale gegenueber rein globaler Skalierung?
4. Wann ueberwiegt Scheduling- und Index-Overhead den Vorteil?
5. Wie stark helfen Viewport- und Interaktionssignale im Web?
6. Welche Codierungsbasis eignet sich fuer Refinements am besten?

15 Schlussfolgerung

REIST IMAGE formuliert Bilduebertragung als fortlaufende Entscheidung: *Welche naechsten Bytes erhoehen den visuellen Informationsgehalt am staerksten?* Layout und Vektorskizze stellen frueh Komposition und Geometrie bereit; ein kleiner Raster-Basislayer und skalierbare globale sowie regionale Refinements investieren weitere Bandbreite dort, wo sie voraussichtlich den groessten Nutzen erzeugt. Der Ansatz ist bewusst als testbare Hypothese formuliert. Erst reproduzierbare Messreihen gegen etablierte progressive Verfahren koennen zeigen, ob daraus ein eigenstaendiger praktischer Beitrag entsteht.

Status und Reproduzierbarkeit

Dieses Dokument beschreibt einen implementierten Forschungsprototypen und erste interne Messungen. Diese Messungen sind noch keine reproduzierbare vergleichende Evaluation von REIST IMAGE. Jede spaetere Leistungsbehauptung soll mit Version, Datensatz, Netzprofil, Metrik und Referenzverfahren dokumentiert werden.

Literatur

- [1] G. K. Wallace, “The JPEG Still Picture Compression Standard,” *Communications of the ACM*, 34(4), 30–44, 1991. doi:10.1145/103085.103089.
- [2] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice*. Kluwer/Springer, 2002. doi:10.1007/978-1-4615-0799-4.
- [3] JPEG Committee, “JPEG 2000 Documentation,” ISO/IEC JTC 1/SC 29/WG 1, <https://jpeg.org/jpeg2000/documentation.html>.
- [4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Trans. Image Processing*, 13(4), 600–612, 2004.
- [5] K. Oku and L. Pardue, “Extensible Prioritization Scheme for HTTP,” RFC 9218, IETF, 2022. <https://www.rfc-editor.org/rfc/rfc9218.html>.