

THÈSE DE DOCTORAT DE

NANTES UNIVERSITÉ

ÉCOLE DOCTORALE N° 641
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Informatique*

Par

Andréas PASTOR

Local Visual Perceptual Differences in Video: Psychophysics methods, computational prediction, and application to Perceptual Encoding in Open Video Codecs

Thèse présentée et soutenue à Polytech Nantes, le 26 août 2024
Unité de recherche : UMR6004–LS2N

Rapporteurs avant soutenance :

Mathias WIEN Priv. Doz. Dr.-Ing., RWTH Aachen University, Allemagne
Aladine CHETOUANI Maître de Conférences, HDR, Université d'Orléans, France

Composition du Jury :

Président :	Aline ROUMY	Directrice de Recherche, INRIA, Rennes, France
Examineurs :	Mathias WIEN	Professeur, RWTH Aachen University, Allemagne
	Aladine CHETOUANI	Maître de Conférences, HDR, Université d'Orléans, France
	Luce MORIN	Professeure des universités, INSA Rennes, France
	Giuseppe VALENZISE	Chargé de Recherche, HDR, CNRS, Paris Saclay, France
Dir. de thèse :	Patrick LE CALLET	Professeur des universités, Nantes Université, France

Invité(s) :

Lukáš KRASULA Senior Research Scientist - Video Quality Metrics, Netflix Inc., Los Gatos, USA

Titre : Différences Perceptuelles Visuelles Locales dans les séquences d'images : méthodes Psychophysiques, prédiction algorithmique, et application à l'Encodage Perceptuel dans les Codecs Vidéo Ouverts

Mot clés : Perception humaine, méthodologies d'évaluation de la qualité, métriques, codecs

Résumé : Développer une métrique qui fonctionne à un horizon spatio-temporel localisé est hautement souhaitable pour piloter les décisions de codage. Si VMAF est très efficace dans certains cas d'usage, il est développé et validé pour de larges horizons spatio-temporels, présidant généralement à la qualité d'une sous-région/bloc d'une vidéo entière. Agir localement dans le codage vidéo offre d'avantages de degrés de liberté pour l'optimisation de la compression. Un codage perceptuel conduit à une qualité vidéo améliorée et à une meilleure efficacité de transmission. Ce travail présente d'abord les Unités Perceptuelles (PU). Ce nouveau concept explore de petits horizons spatio-temporels inspirés de

comment le système visuel humain explore le contenu avec le regard et des saccades. Nous étudions de nombreux paradigmes et méthodologies d'évaluation de la qualité pour collecter des données subjectives en vidéo à l'échelle PU vers un ensemble de données à grande échelle d'évaluation des distorsions locales. Ce vaste ensemble de données sert à explorer la pertinence de caractéristiques basées apprentissage machine ou de substrats de la perception visuel humaine pour entraîner une métrique frugale pour la prédiction des Différences de Perception (PD). La dernière contribution est l'intégration de cette métrique de qualité locale dans l'encodeur AV1 pour un meilleur encodage perceptuel.

Title: Local Visual Perceptual Differences in Video: Psychophysics methods, computational prediction, and application to Perceptual Encoding in Open Video Codecs

Keywords: Human perception, methodologies for quality assessment, metrics, video codecs

Abstract: Developing a metric that operates at a localized spatio-temporal horizon is highly desirable to drive video encoding decisions. While VMAF is very efficient in some use cases, it is developed and validated for large spatiotemporal horizons, generally presiding over the quality of a sub-region/block of an entire video. Acting locally in video coding offers more degrees of freedom for perceptual optimization. This perceptual encoding can lead to improved video quality and better transmission efficiency. This thesis first introduces Perceptual Units (PU). This novel concept explore a spatio-temporal range inspired from how the Human Visual System

explores content with gazes and saccades. We investigate numerous quality assessment paradigms and methodologies to collect subjective data in video at the PU scale towards a large-scale dataset of local distortions evaluation. These investigations led to innovations in tools to collect data efficiently, be robust in crowdsourcing scenarios to outliers, and report on the discriminability of a subjective test. This large dataset serves to explore the relevance of handcrafted and deep features to train a lightweight metric for Perceptual Differences (PD) prediction. Final contribution is integrating this local quality metric in the AV1 encoder for better perceptual encoding.