

NORME
INTERNATIONALE

ISO/CEI
13818-2

Première édition
1996-05-15

**Technologies de l'information — Codage
générique des images animées et du son
associé: Données vidéo**

*Information technology — Generic coding of moving pictures and
associated audio information: Video*

ISO/IEC 13818-2:1996

<https://standards.iteh.ai/catalog/standards/sist/e057f082-fd83-43ae-8bda-8feb6585d4ae/iso-iec-13818-2-1996>



Numéro de référence
ISO/CEI 13818-2:1996(F)

Sommaire

	<i>Page</i>	
1	Domaine d'application.....	1
2	Références normatives.....	1
3	Définitions.....	2
4	Abréviations et symboles.....	8
4.1	Opérateurs arithmétiques.....	8
4.2	Opérateurs logiques.....	9
4.3	Opérateurs relationnels.....	9
4.4	Opérateurs binaires.....	9
4.5	Affectation.....	9
4.6	Mnémoniques.....	10
4.7	Constantes.....	10
5	Conventions.....	10
5.1	Méthode de description de la syntaxe du flux binaire.....	10
5.2	Définition des fonctions.....	11
5.3	Valeur réservée, valeur interdite et bit marqueur.....	12
5.4	Précision arithmétique.....	12
6	Syntaxe et sémantique du flux binaire de données vidéo.....	12
6.1	Structure des données vidéo codées.....	12
6.2	Syntaxe du flux binaire de données vidéo codées.....	24
6.3	Sémantique du flux binaire de données vidéo codées.....	40
7	Processus de décodage des données vidéo.....	66
7.1	Structures syntaxiques supérieures.....	66
7.2	Décodage à longueur variable.....	67
7.3	Balayage inverse des coefficients.....	70
7.4	Quantification inverse.....	71
7.5	Transformation DCT inverse.....	76
7.6	Compensation de mouvement.....	76

© ISO/CEI 1996

Droits de reproduction réservés. Sauf prescription différente, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de l'éditeur.

ISO/CEI Copyright Office • Case postale 56 • CH-1211 Genève 20 • Suisse

Imprimé en Suisse

7.7	Echelonnabilité spatiale	91
7.8	Echelonnabilité SNR.....	104
7.9	Echelonnabilité temporelle	109
7.10	Subdivision des données	113
7.11	Echelonnabilité hybride	114
7.12	Sortie du processus de décodage.....	115
8	Profils et niveaux.....	119
8.1	Compatibilité avec ISO/CEI 11172-2	120
8.2	Relation entre profils définis.....	120
8.3	Relation entre niveaux définis.....	122
8.4	Couches échelonnables	123
8.5	Valeurs paramétriques pour profils, niveaux et couches définis.....	126
Annexe A	– Transformation en cosinus discrète.....	130
Annexe B	– Tables des codes à longueur variable	131
B.1	Adressage des macroblocs	131
B.2	Type de macrobloc.....	132
B.3	Structure des macroblocs	137
B.4	Vecteurs de mouvement.....	138
B.5	Coefficients DCT.....	139
Annexe C	– Vérificateur de mémoire vidéo.....	147
Annexe D	– Caractéristiques supportées par l'algorithme.....	152
D.1	Vue d'ensemble	152
D.2	Formats vidéo.....	152
D.3	Qualité d'image	153
D.4	Contrôle du débit binaire	153
D.5	Mode à faible délai.....	154
D.6	Accès aléatoire/interconnexion des canaux	154
D.7	Echelonnabilité	154
D.8	Compatibilité.....	162
D.9	Différences entre la présente Spécification et ISO/CEI 11172-2.....	163
D.10	Complexité.....	165
D.11	Edition des flux binaires codés	165
D.12	Modes d'enrichissement	166
D.13	Robustesse aux erreurs.....	167
D.14	Séquences concaténées	176
Annexe E	– Restrictions de profil et de niveau	177
E.1	Restrictions applicables aux éléments syntaxiques dans les profils.....	177
E.2	Combinaisons de couches autorisées	189
Annexe F	– Bibliographie	211

Avant-propos

L'ISO (Organisation internationale de normalisation) et la CEI (Commission électrotechnique internationale) forment ensemble un système consacré à la normalisation internationale considérée comme un tout. Les organismes nationaux membres de l'ISO ou de la CEI participent au développement de Normes internationales par l'intermédiaire des comités techniques créés par l'organisation concernée afin de s'occuper des différents domaines particuliers de l'activité technique. Les comités techniques de l'ISO et de la CEI collaborent dans des domaines d'intérêt commun. D'autres organisations internationales, gouvernementales ou non gouvernementales, en liaison avec l'ISO et la CEI participent également aux travaux.

Dans le domaine des technologies de l'information, l'ISO et la CEI ont créé un comité technique mixte, l'ISO/CEI JTC 1. Les projets de Normes internationales adoptés par le comité technique mixte sont soumis aux organismes nationaux pour approbation, avant leur acceptation comme Normes internationales. Les Normes internationales sont approuvées conformément aux procédures qui requièrent l'approbation de 75 % au moins des organismes nationaux votants.

La Norme internationale ISO/CEI 13818-2 a été élaborée par le comité technique mixte ISO/CEI JTC 1, *Technologies de l'information*, sous-comité SC 29, *Codage du son, de l'image, de l'information multimédia et hypermédia*, en collaboration avec l'UIT-T. Le texte identique est publié en tant que Recommandation UIT-T H.262.

L'ISO/CEI 13818 comprend les parties suivantes, présentées sous le titre général *Technologies de l'information — Codage générique des images animées et du son associé*:

- *Partie 1: Systèmes*
- *Partie 2: Données vidéo*
- *Partie 3: Son*
- *Partie 4: Essais de conformité*
- *Partie 6: Extensions pour DSM-CC*
- *Partie 9: Extension pour interface temps réel pour systèmes décodeurs*

Les annexes A à C font partie intégrante de la présente partie de l'ISO/CEI 13818. Les annexes D à F sont données uniquement à titre d'information.

Introduction

Intro. 1 Objet

La présente Partie de cette Spécification a été élaborée en réponse au besoin croissant d'une méthode de codage générique des images animées et du son associé, pour diverses applications telles que les supports numériques d'enregistrement, la télédiffusion et la communication. L'utilisation de la présente Spécification implique que les données vidéo d'images animées peuvent être manipulées sous forme de données informatiques, stockées sur divers supports d'enregistrement, être émises et reçues au moyen des réseaux existants et futurs, puis distribuées sur les canaux de télédiffusion existants et futurs.

Intro. 2 Application

Les applications de la présente Spécification couvrent différents domaines tels que:

CaTV	Télévision par câble, distribuée par câbles à fibres optiques, à conducteurs métalliques, etc. (<i>cable TV</i>)
CDAD	Distribution audionumérique par câble (<i>Cable Digital Audio Distribution</i>)
DSB	Diffusion audionumérique (de Terre et par satellite) (<i>Digital Sound Broadcasting</i>)
DTTB	Télédiffusion numérique par voie hertzienne de Terre (<i>Digital Terrestrial Television Broadcasting</i>)
EC	Cinéma électronique (<i>Electronic Cinema</i>)
HTT	Spectacle télédiffusé à domicile (<i>Home Television Theatre</i>)
IPC	Communications interpersonnelles (<i>Interpersonal Communications</i>) (visioconférence, visio- phone, etc.)
ISM	Supports d'enregistrements interactifs (<i>Interactive Storage Media</i>) (disques optiques, etc.)
JET	Journalisme Electronique Télévisuel, y compris le Reportage d'Actualités par Satellite (RAS)
MMM	Messagerie multimédia (<i>Multimedia Mailing</i>)
NCA	Nouvelles et actualités (<i>News and Current Affairs</i>)
NDB	Base de données sur réseau (via ATM, etc.) (<i>Networked Database</i>)
RVS	Télévidéosurveillance (<i>Remote Video Surveillance</i>)
SFS	Service Fixe par Satellite (par exemple, vers des têtes de réseau)
SRS	Service de Radiodiffusion par Satellite (à domicile)
SSM	Support d'enregistrement séquentiel (magnétoscopes numériques, etc.) (<i>Serial Storage Media</i>)

Intro. 3 Profils et niveaux

La présente Spécification a une vocation générique, en ce sens qu'elle vise une large gamme d'applications, de débits, de résolutions, de qualités et de services. Les applications devraient couvrir entre autres les supports d'enregistrement numérique, la télédiffusion et les communications. Au cours de l'élaboration de la présente Spécification, diverses exigences ont été prises en considération sur la base d'applications typiques. Les éléments algorithmiques nécessaires ont été mis au point et ont été intégrés dans une syntaxe unique. C'est pourquoi la présente Spécification facilitera l'échange de flux binaires entre applications différentes.

Compte tenu cependant des impératifs d'application pratique de la syntaxe générale décrite dans la présente Spécification, un nombre limité de modules de cette syntaxe sont également stipulés en tant que «profils» et «niveaux». Ces termes, et leurs analogues, sont définis formellement à l'article 3.

Un «profil» est un sous-ensemble défini de la syntaxe générale du flux binaire, elle-même définie par la présente Spécification. Dans le cadre des limites imposées par la syntaxe d'un profil donné, il est toujours possible de prescrire une très grande étendue de variation de performance pour les codeurs et les décodeurs, en fonction des valeurs attribuées aux paramètres du flux binaire. Il est par exemple possible de spécifier des images de dimensions allant jusqu'à environ 2^{14} pels en largeur par 2^{14} lignes en hauteur. Pour le moment, il n'est ni pratique ni économique de mettre en œuvre un décodeur capable de traiter toutes les tailles d'image possibles.

Pour résoudre ce problème, des «niveaux» sont définis dans chaque profil. Un niveau est défini comme étant un ensemble de contraintes imposées sur des paramètres dans le flux binaire. Ces contraintes peuvent être de simples limites numériques. Elles peuvent également prendre la forme de limites imposées à des combinaisons arithmétiques des paramètres (par exemple, largeur d'image fois la hauteur d'image fois la fréquence image).

Les flux binaires conformes à la présente Spécification font appel à une syntaxe commune. Afin d'obtenir un sous-ensemble de la syntaxe complète, on inclut dans le flux binaire des fanions et des paramètres qui signalent la présence ou l'absence d'éléments syntaxiques devant apparaître ultérieurement dans le flux binaire. Pour spécifier des contraintes syntaxiques (et donc définir un profil), il suffit donc de délimiter les valeurs de ces fanions et paramètres, spécifiant la présence d'éléments syntaxiques ultérieurs.

Intro. 4 Syntaxe échelonnable et syntaxe non échelonnable

On peut subdiviser la syntaxe complète en deux catégories principales: l'une est la syntaxe non échelonnable, qui est structurée comme un surensemble de la syntaxe définie dans ISO/CEI 11172-2. La principale caractéristique de la syntaxe non échelonnable est la présence d'outils supplémentaires de compression pour les signaux vidéo entrelacés. La deuxième catégorie est la syntaxe échelonnable, dont la principale caractéristique est de permettre la reconstruction de données vidéo utiles à partir de segments d'un flux binaire total. Le processus consiste à structurer le flux total en deux couches différentes ou plus, en commençant par une couche de base autonome et en ajoutant un certain nombre de couches d'amélioration. La couche de base peut utiliser la syntaxe non échelonnable ou, en certaines occurrences, utiliser une syntaxe conforme à ISO/CEI 11172-2.

Intro. 4.1 Vue d'ensemble de la syntaxe non échelonnable

La représentation codée qui est définie dans la syntaxe non échelonnable réalise un taux de compression élevé tout en préservant une bonne qualité d'image. L'algorithme n'est pas sans perte car les valeurs exactes des pels ne sont pas conservées au cours du codage. L'obtention d'une bonne qualité iconographique aux débits binaires en cause exige un taux de compression très élevé, ce qui n'est pas réalisable avec le seul codage intra-image. La nécessité d'un accès aléatoire est toutefois mieux prise en compte avec un pur codage intra-image. Le choix entre les techniques est fondé sur le besoin de trouver un compromis entre une qualité d'image élevée, un taux de compression élevé et la nécessité de permettre un accès aléatoire au flux binaire codé.

Un certain nombre de techniques sont utilisées pour obtenir un taux de compression élevé. L'algorithme utilise d'abord une compensation de mouvement par blocs, ce qui permet de diminuer la redondance temporelle. Cette compensation de mouvement est utilisée aussi bien pour la prédiction causale de l'image actuelle à partir d'une image précédente que pour la prédiction non causale (interpolative) à partir d'images précédentes et futures. Les vecteurs de mouvement sont définis pour chaque région de 16 pels × 16 lignes de l'image. Le signal différentiel, c'est-à-dire l'erreur de prédiction, est encore comprimé au moyen de la transformée en cosinus discrète (DCT) afin d'en éliminer les corrélations spatiales avant sa quantification au cours d'un processus irréversible qui rejette les informations moins importantes. Finalement, les vecteurs de mouvement sont combinés avec les informations résiduelles de transformation DCT puis codés au moyen de mots à longueur variable.

Intro. 4.1.1 Traitement temporel

En raison des exigences contradictoires de l'accès aléatoire et d'une compression très efficace, trois principaux types d'image seront définis. Les images à codage intratrame ou images intra (images I) sont codées sans référence à d'autres images. Elles fournissent des points d'accès à la séquence codée, à partir desquels le décodage peut commencer; mais elles ne sont codées qu'à un taux de compression modeste. Les images à codage prédictif ou images prédites (images P) sont codées plus efficacement, avec une prédiction basée sur la compensation de mouvement à partir d'une précédente image intra ou prédite. Les images P sont généralement utilisées comme références pour la prédiction des images suivantes. Les images codées par prédiction bidirectionnelle ou images bidirectionnelles (images B) offrent le taux de compression le plus élevé mais nécessitent la présence d'images de référence aussi bien antérieures que postérieures pour effectuer la compensation de mouvement. Ces images bidirectionnelles ne sont jamais utilisées comme références pour la prédiction (sauf dans le cas où l'image résultante est utilisée comme référence dans une couche d'amélioration spatialement échelonnable). L'organisation de ces trois types d'image dans une séquence est très souple. Le choix en est laissé au codeur et dépendra des exigences de l'application. La Figure Intro. 1 montre la relation entre les trois différents types d'image.

Intro. 4.1.2 Codage d'images vidéo entrelacées

Chaque image d'une vidéo entrelacée est formée de deux trames, séparées par un signal de synchronisation de trame. Selon la présente Spécification, les deux trames d'une image complète peuvent être codées soit comme une seule image soit comme deux images. Le codage image (ou «bi-trame») ou le codage trame (ou monotrème) peut être sélectionné dynamiquement, image par image. Le codage image est normalement préféré lorsque la scène vidéo contient un nombre important de détails avec peu de mouvements. Le codage trame, dans lequel la deuxième trame peut être prédite d'après la première, donne de meilleurs résultats lorsque les mouvements sont rapides.

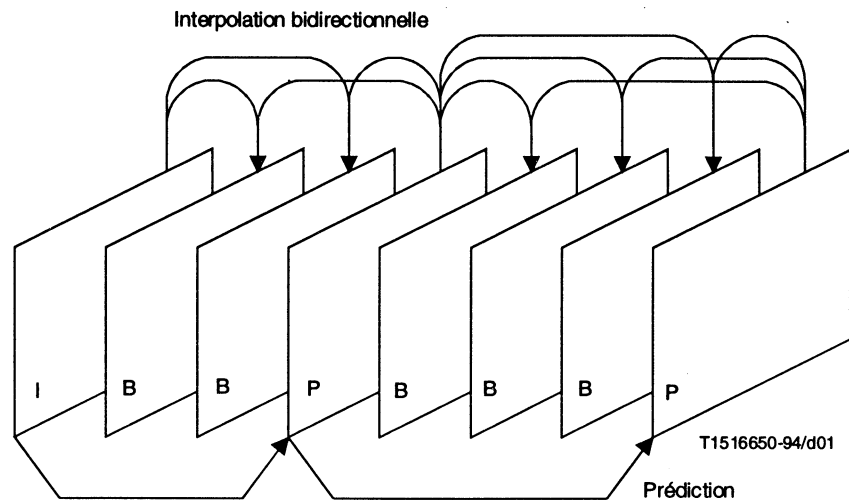


Figure Intro. 1 – Exemple de structure temporelle d'images

Intro. 4.1.3 Représentation du mouvement – Macroblocs

Comme dans ISO/CEI 11172-2, le choix de 16×16 macroblocs pour l'unité de compensation de mouvement est le résultat d'un compromis entre le gain de codage obtenu en utilisant les informations de mouvement et la servitude binaire due à leur représentation. Chaque macrobloc peut être temporellement prédit, selon une parmi plusieurs méthodes différentes. Par exemple, en codage image, la prédiction de la précédente image peut elle-même se fonder sur les deux trames de cette image ou sur une seule. Selon le type de macrobloc, l'information du vecteur de mouvement, et toutes autres informations associées, est codée avec un signal d'erreur de prédiction dans chaque macrobloc. Les vecteurs de mouvement sont codés différemment tout en respectant les derniers vecteurs de mouvement codés au moyen de mots à longueur variable. On peut programmer image par image la longueur maximale des vecteurs de mouvement, de manière que les applications les plus exigeantes puissent être mises en œuvre sans compromettre la performance du système dans des situations plus courantes.

Il appartient au codeur de calculer les vecteurs de mouvement appropriés. La présente Spécification ne spécifie pas la façon dont il convient d'effectuer ce calcul.

Intro. 4.1.4 Réduction de la redondance spatiale

Les signaux des images originales et les signaux d'erreur de prédiction possèdent, les uns comme les autres, un degré élevé de redondance spatiale. La présente Spécification fait appel à une méthode de filtrage par transformée DCT sur des blocs, avec quantification pondérée en termes de perception visuelle et codage des longueurs de séquence. Après une prédiction compensée en mouvement ou une interpolation, l'image résiduelle est subdivisée en 8×8 blocs. Ces blocs sont convertis dans le domaine DCT, où ils sont ensuite pondérés avant d'être quantifiés. A l'issue du processus de quantification, de nombreux coefficients ont une valeur nulle. On fait alors appel à un codage à longueur variable sur un tableau à 2 dimensions afin de coder efficacement tous les coefficients.

Intro. 4.1.5 Formats de chrominance

En plus du format 4:2:0 qui est retenu dans ISO/CEI 11172-2, la présente Spécification supporte les formats de chrominance 4:2:2 et 4:4:4.

Intro. 4.2 Extensions échelonnables

Les outils d'échelonnabilité figurant dans la présente Spécification sont conçus pour gérer des applications pouvant supporter plusieurs couches vidéo. Entre autres domaines d'application notables, on citera les télécommunications vidéo, les communications vidéo sur réseaux à mode de transfert asynchrone (ATM), l'interfonctionnement des standards vidéo, la hiérarchisation des services vidéo selon diverses résolutions spatiales, temporelles et qualitatives, la TVHD à TV

imbriquée, les systèmes autorisant la migration vers une TVHD à résolution temporelle plus élevée, etc. Une solution simple, en termes de vidéo échelonnable, est la technique de diffusion simultanée qui est fondée sur la transmission ou l'enregistrement, après lecture du support, de multiples sources vidéo codées indépendamment les unes des autres; mais une autre solution, plus efficace, consiste à effectuer un codage échelonnable des données vidéo, dans lequel la largeur de bande attribuée à une reproduction donnée du support vidéo peut être partiellement réutilisée pour le codage de la reproduction vidéo suivante. Dans un codage vidéo échelonnable, on part du principe que, pour chaque type de flux binaire codé, des décodeurs présentant divers degrés de complexité peuvent décoder et afficher des données vidéo codées, lues de manière appropriée. Un codeur échelonnable est susceptible d'être plus complexe qu'un codeur à une seule couche. La présente Recommandation / Norme internationale distingue toutefois plusieurs modes d'échelonnabilité, s'adressant à des applications ne se recoupant pas, de complexité correspondante. Les principaux outils d'échelonnabilité offerts sont les suivants:

- subdivision des données;
- échelonnabilité SNR;
- échelonnabilité spatiale; et
- échelonnabilité temporelle.

Par ailleurs, des combinaisons de ces outils d'échelonnabilité de base sont également possibles: de telles combinaisons seront appelées *échelonnabilité hybride*. Dans le cas de l'échelonnabilité de base, on peut avoir deux couches de flux vidéo: la *couche inférieure* et la *couche d'amélioration*; alors que, dans le cas de l'échelonnabilité hybride, on peut avoir jusqu'à trois couches de données. Les Tableaux Intro. 1 à Intro. 3 présentent quelques exemples d'application selon divers degrés d'échelonnabilité.

iTeh STANDARD PREVIEW

Tableau Intro. 1 – Applications de l'échelonnabilité SNR

Couche inférieure	Couche d'amélioration	Application
Recommandation UIT-R BT.601	Même résolution et même format que la couche inférieure	Service à deux niveaux de qualité pour la télévision conventionnelle
Haute définition	Même résolution et même format que la couche inférieure	Service à deux niveaux de qualité pour la TVHD
Haute définition en 4:2:0	Diffusion simultanée en format chromatique 4:2:2	Production/distribution vidéo

Tableau Intro. 2 – Applications de l'échelonnabilité spatiale

Couche de base	Couche d'amélioration	Application
Balayage progressif (30 Hz)	Balayage progressif (30 Hz)	Compatibilité ou échelonnabilité CIF/SCIF
Balayage entrelacé (30 Hz)	Balayage entrelacé (30 Hz)	Echelonnabilité TVHD/SDTV
Balayage progressif (30 Hz)	Balayage entrelacé (30 Hz)	Compatibilité avec l'ISO/CEI 11172-2 ou avec la présente Spécification
Balayage entrelacé (30 Hz)	Balayage progressif (60 Hz)	Migration vers TVHD à haute résolution temporelle et balayage progressif

Tableau Intro. 3 – Applications de l'échelonnabilité temporelle

Couche de base	Couche d'amélioration	Couches supérieures	Application
Balayage progressif (30 Hz)	Balayage progressif (30 Hz)	Balayage progressif (60 Hz)	Migration vers TVHD à haute résolution temporelle et balayage progressif
Balayage entrelacé (30 Hz)	Balayage entrelacé (30 Hz)	Balayage progressif (60 Hz)	Migration vers TVHD à haute résolution temporelle et balayage progressif

Intro. 4.2.1 Extension à échelonnabilité spatiale

L'échelonnabilité spatiale est un outil destiné à être utilisé dans des applications vidéo telles que les télécommunications, l'interfonctionnement de standards vidéo, la consultation de bases de données vidéo, l'interfonctionnement TV-TVHD, etc., c'est-à-dire dans des systèmes vidéo dont la principale caractéristique commune est l'exigence d'au moins deux couches de résolution spatiale. L'échelonnabilité spatiale implique la production, à partir d'une seule source vidéo, de deux couches de résolution spatiale telles que la couche inférieure soit codée indépendamment pour fournir la couche de base de la résolution spatiale et que la couche d'amélioration utilise cette couche inférieure comme base d'interpolation spatiale pour apporter la pleine résolution spatiale de la source d'entrée vidéo. La couche de base inférieure et la couche d'amélioration peuvent soit utiliser toutes les deux les outils de codage décrits dans la présente Spécification, ou bien utiliser les outils de l'ISO/CEI 11172-2 pour la couche inférieure et ceux de la présente Spécification pour la couche supérieure d'amélioration. Cette dernière variante offre l'avantage supplémentaire de faciliter l'interfonctionnement entre normes de codage vidéo. En outre, l'échelonnabilité spatiale offre la possibilité de choisir le format vidéo à employer dans chaque couche. Elle permet aussi d'assurer une meilleure robustesse aux erreurs de transmission en acheminant les données plus importantes de la couche inférieure par un canal présentant de meilleures caractéristiques en termes de protection contre les erreurs, tandis que les données moins critiques de la couche d'amélioration peuvent être acheminées par un canal présentant une protection d'erreur de qualité moindre.

[https://standards.iteh.ai/catalog/standards/sist/e057f082-fd83-43ae-8bda-8feb6585d4ae/iso-iec-](https://standards.iteh.ai/catalog/standards/sist/e057f082-fd83-43ae-8bda-8feb6585d4ae/iso-iec-13818-2-1996)

Intro. 4.2.2 Extension à échelonnabilité SNR

L'échelonnabilité SNR (*Signal-Noise Ratio*) (rapport signal sur bruit) est un outil destiné à être utilisé dans des applications vidéo telles que les télécommunications, les services vidéo de diverses qualités, la TV conventionnelle et la TVHD, c'est-à-dire dans des systèmes vidéo dont la principale caractéristique commune est l'exigence d'au moins deux couches de qualité vidéo. L'échelonnabilité SNR implique la production, à partir d'une seule source vidéo, de deux couches vidéo de même résolution spatiale, telles que la couche inférieure soit codée indépendamment pour fournir la couche de base de la qualité vidéo et que la couche d'amélioration soit codée de manière à renforcer cette couche inférieure. Une fois ajoutée à la couche inférieure, la couche d'amélioration apporte une meilleure qualité de reproduction de la source vidéo d'entrée. La couche de base inférieure et la couche d'amélioration peuvent soit utiliser toutes les deux les outils de codage décrits dans la présente Spécification, ou bien utiliser les outils de l'ISO/CEI 11172-2 pour la couche inférieure et ceux de la présente Spécification pour la couche supérieure. L'échelonnabilité SNR permet aussi d'assurer une meilleure robustesse aux erreurs de transmission en acheminant les données plus importantes de la couche inférieure par un canal présentant de meilleures caractéristiques en termes de protection contre les erreurs, tandis que les données moins critiques de la couche d'amélioration peuvent être acheminées par un canal présentant une protection d'erreur de qualité moindre.

Intro. 4.2.3 Extension à échelonnabilité temporelle

L'échelonnabilité temporelle est un outil destiné à être utilisé dans une gamme d'applications vidéo diverses, allant des télécommunications à la TVHD, pour lesquelles il peut être nécessaire d'assurer une migration vers des systèmes offrant une résolution temporelle supérieure à celle d'autres systèmes analogues. Souvent, les systèmes vidéo à faible résolution temporelle pourront être les systèmes existants actuellement ou les systèmes moins coûteux des premières générations, l'objectif étant d'introduire progressivement des systèmes plus évolués. L'échelonnabilité temporelle implique la stratification des images vidéo: alors que la couche inférieure est codée indépendamment pour fournir la base de temps principale, la couche d'amélioration subit un codage temporel prédictif par rapport à la couche inférieure. Une fois

décodées et démultiplexées dans le temps, ces deux couches fournissent la pleine résolution temporelle de la source vidéo. Les systèmes vidéo à faible résolution temporelle ne peuvent décoder que la couche inférieure pour fournir la résolution temporelle de base, tandis que les futurs systèmes plus évolués pourront décoder les deux couches et fournir des signaux de haute résolution temporelle tout en conservant l'interfonctionnement avec les systèmes vidéo des générations antérieures. Un avantage complémentaire de l'échelonnabilité temporelle est d'offrir une certaine élasticité aux erreurs de transmission étant donné que les données de la couche inférieure, qui sont les plus importantes, peuvent être acheminées par un canal présentant de meilleures caractéristiques en termes de protection contre les erreurs, tandis que les données moins critiques de la couche d'amélioration peuvent être acheminées par un canal présentant une moindre qualité en termes de protection contre les erreurs.

Intro. 4.2.4 Extension vers la subdivision des données

La subdivision des données est un outil destiné à être utilisé lorsque deux canaux sont disponibles pour la transmission et/ou pour l'enregistrement d'un flux binaire vidéo, comme ce peut être le cas dans les réseaux ATM, en radiodiffusion de Terre, avec des supports magnétiques, etc. Le flux binaire est subdivisé entre ces canaux de manière que ses parties les plus critiques (comme les en-têtes, les vecteurs de mouvement, les coefficients DCT) soient transmises dans le canal offrant les meilleures caractéristiques en termes de protection contre les erreurs et que les données moins critiques (telles que les coefficients DCT d'ordre supérieur) soient transmises dans le canal présentant une moindre qualité en termes de protection contre les erreurs. Ce procédé permet de minimiser les erreurs de type dégradation introduite dans le canal car les parties critiques du flux binaire sont dans un canal mieux protégé. Aucune donnée de ces deux canaux ne peut être traitée par un décodeur non destiné au décodage de flux binaires à subdivision des données.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[ISO/IEC 13818-2:1996](https://standards.iteh.ai/catalog/standards/sist/e057f082-fd83-43ae-8bda-8feb6585d4ae/iso-iec-13818-2-1996)

<https://standards.iteh.ai/catalog/standards/sist/e057f082-fd83-43ae-8bda-8feb6585d4ae/iso-iec-13818-2-1996>

NORME INTERNATIONALE

RECOMMANDATION UIT-T

TECHNOLOGIES DE L'INFORMATION – CODAGE GÉNÉRIQUE DES IMAGES ANIMÉES ET DU SON ASSOCIÉ: DONNÉES VIDÉO

1 Domaine d'application

La présente Recommandation | Norme internationale spécifie la représentation codée des informations d'image pour supports d'enregistrement numérique et vidéocommunications, ainsi que le processus de décodage correspondant. Cette représentation est compatible avec la transmission à débit binaire constant, la transmission à débit binaire variable, l'accès aléatoire, l'interconnexion des canaux, le décodage échelonné, l'édition du flux binaire ainsi que des fonctions spéciales telles que la lecture rapide en avant, la lecture rapide en arrière, le ralenti, la pause et les arrêts sur image. La présente Recommandation | Norme internationale est compatible en aval avec ISO/CEI 11172-2 et compatible aussi bien en amont qu'en aval avec les formats de télévision à définition améliorée (EDTV), de télévision à haute définition (TVHD) et de télévision conventionnelle (SDTV).

La présente Recommandation | Norme internationale est principalement applicable aux supports d'enregistrement numérique, à la vidéodiffusion et aux vidéocommunications. Les supports d'enregistrement peuvent être reliés au décodeur directement ou par l'intermédiaire de moyens de communication tels que des bus d'interconnexion, des réseaux locaux ou des liaisons de télécommunication.

2 Références normatives

Les Recommandations UIT-T et les Normes internationales suivantes contiennent des dispositions qui, par suite de la référence qui y est faite, constituent des dispositions valables pour la présente Recommandation | Norme internationale. Au moment de la publication, les éditions indiquées étaient en vigueur. Toute Recommandation ou Norme internationale est sujette à révision et les parties prenantes aux accords fondés sur la présente Recommandation | Norme internationale sont invitées à rechercher la possibilité d'appliquer les éditions les plus récentes des Recommandations et Normes indiquées ci-après. Les membres de la CEI et de l'ISO possèdent le registre des Normes internationales en vigueur. Le Bureau de la normalisation des télécommunications (TSB) de l'UIT tient à jour une liste des Recommandations UIT-T en vigueur.

- Recommandations et Rapports du CCIR, 1990, XVII^e Assemblée plénière, Düsseldorf, 1990 – Volume XI – Partie 1 – Service de radiodiffusion (télévision) – Recommandation UIT-R B.T.601-3, *Paramètres de codage de télévision numérique pour studios.*
- Volumes X et XI (Partie 3) du CCIR – Recommandation UIT-R BR.648, *Enregistrement numérique des signaux audio.*
- Volumes X et XI (Partie 3) du CCIR – Rapport UIT-R 955-2, *Radiodiffusion sonore par satellite pour récepteurs portatifs et récepteurs dans des véhicules automobiles.*
- ISO/CEI 11172-1:1993, *Technologies de l'information – Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbit/s – Partie 1: Systèmes.*
- ISO/CEI 11172-2:1993, *Technologies de l'information – Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbit/s – Partie 2: Vidéo.*
- ISO/CEI 11172-3:1993, *Technologies de l'information – Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbit/s – Partie 3: Audio.*
- IEEE Standard Specifications for the Implementations of 8 by 8 Inverse Discrete Cosine Transform (Spécifications de norme IEEE pour les applications de la DCT inverse à des blocs de 8 × 8 éléments), IEEE Std. 1180-1990, 6 décembre 1990.

- Publication CEI 908:1987, *Système audionumérique à disque compact*.
- Publication CEI 461:1986, *Code temporel de commande pour les magnétoscopes*.
- Recommandation UIT-T H.261 (1993), *Codec vidéo pour services audiovisuels à $p \times 64$ kbit/s*.
- Recommandation T.81 du CCITT (1992) (JPEG) | ISO/CEI 10918-1:1994, *Technologie de l'information – Compression numérique et codage des images fixes de nature photographique – Prescriptions et lignes directrices*.

3 Définitions

Pour les besoins de la présente Recommandation | Norme internationale, les définitions suivantes s'appliquent.

- 3.1 coefficient AC:** tout coefficient DCT pour lequel la fréquence n'est pas nulle dans une dimension ou dans les deux dimensions.
- 3.2 image non décodée:** image codée qui provoquerait le sous-remplissage de la mémoire tampon VBV selon la définition donnée en C.7. De telles images ne peuvent apparaître qu'à l'intérieur de séquences dont le champ `low_delay` est mis à la valeur '1'. L'expression «image sautée» désigne parfois le même concept.
- 3.3 image B monotrame:** image B structurée en trame unique.
- 3.4 image B bi-trame:** image B structurée en deux trames.
- 3.5 image B; image à codage prédictif bidirectionnel:** image qui est codée au moyen de la prédiction utilisant la compensation de mouvement à partir de précédentes et/ou futures trames de référence uniques ou doubles.
- 3.6 compatibilité amont:** une nouvelle norme de codage est dite à compatibilité amont avec une ancienne norme de codage si les décodeurs étudiés pour fonctionner avec l'ancienne norme de codage sont en mesure de continuer à fonctionner en décodant tout ou partie du flux binaire produit conformément à la nouvelle norme de codage.
- 3.7 vecteur de mouvement différé:** vecteur de mouvement qui est utilisé pour la compensation de mouvement à partir d'une trame de référence antérieure à ce vecteur dans l'ordre d'affichage.
- 3.8 prédiction différée:** prévision à partir de trames de référence futures.
- 3.9 couche de base:** première couche indépendamment décodable d'une hiérarchie échelonnée.
- 3.10 flux binaire; flux:** série ordonnée d'éléments binaires qui constitue la représentation codée des données.
- 3.11 débit binaire:** débit auquel le flux binaire codé est présenté à l'entrée d'un décodeur à partir d'un support d'enregistrement.
- 3.12 bloc:** matrice de 8 lignes sur 8 colonnes de pels, soit 64 coefficients DCT (originaux, quantifiés ou déquantifiés).
- 3.13 trame inférieure:** l'une des deux trames qui composent une image bi-trame. Chaque ligne d'une trame inférieure est spatialement localisée immédiatement au-dessous de la ligne homologue de la trame supérieure.
- 3.14 limite d'octet:** un bit de flux binaire codé est dit sur une limite d'octet si sa position correspond à un multiple de 8 comptée à partir du premier bit du flux.
- 3.15 octet:** mot de 8 bits.
- 3.16 canal:** support numérique qui conserve ou transporte un flux binaire constitué conformément à la Rec. UIT-T H.262 | ISO/CEI 13818-2.
- 3.17 format de chrominance:** nombre de blocs de chrominance d'un macrobloc.
- 3.18 diffusion jumelée de données de chrominance:** type d'échelonnabilité (qui est un sous-ensemble de l'échelonnabilité SNR), où la (les) couche(s) d'amélioration ne contien(nen)t que des données codées de raffinement pour les coefficients DC et toutes les données pour les coefficients AC, des composantes de chrominance.

- 3.19 chrominance (composante de):** matrice, bloc ou échantillon isolé qui représente un des deux signaux de différence de couleur et qui s'applique aux couleurs primaires selon le mode défini dans le flux binaire. Les symboles utilisés pour les signaux de chrominance sont Cr et Cb.
- 3.20 image bi-trame à codage B:** image B à deux trames ou paire d'images B à trame unique.
- 3.21 image bi-trame codée:** image bi-trame à codage I (intra), P (prédicatif) ou B (bidirectionnel).
- 3.22 image bi-trame à codage I:** image I à deux trames ou paire d'images à trame unique, la première image étant de type I et la deuxième une image de type I ou P.
- 3.23 image bi-trame à codage P:** image P à deux trames ou paire d'images P à trame unique.
- 3.24 image codée:** image composée d'un en-tête d'image, des extensions facultatives qui le suivent immédiatement et des données d'image ultérieures. Une image codée peut être bi-trame ou monotrème.
- 3.25 flux binaire de données vidéo codées:** représentation codée d'une série composée d'une ou de plusieurs images telles que définies dans la Rec. UIT-T H.262 | ISO/CEI 13818-2.
- 3.26 ordre codé:** ordre dans lequel les images sont transmises et décodées. Cet ordre n'est pas nécessairement le même que l'ordre d'affichage.
- 3.27 représentation codée:** élément de données représenté sous forme codée.
- 3.28 paramètres de codage:** ensemble de paramètres définissables par l'utilisateur, qui caractérise un flux binaire de données vidéo. Les flux binaires sont caractérisés par des paramètres de codage. Les décodeurs sont caractérisés par les flux binaires qu'ils sont capables de décoder.
- 3.29 composante:** matrice, bloc ou échantillon isolé dans une des trois matrices (une de luminance et deux de chrominance) constituant une image.
- 3.30 compression:** réduction du nombre d'éléments binaires utilisés pour représenter un élément de données.
- 3.31 données vidéo codées à débit binaire constant:** flux de données vidéo comprimées dont le débit binaire moyen est constant.
- 3.32 débit binaire constant:** exploitation telle que le débit binaire est constant du début à la fin du flux binaire codé.
- 3.33 élément de données:** représentation d'une donnée avant codage et après décodage.
- 3.34 subdivision des données:** méthode consistant à doubler un flux binaire afin d'augmenter l'élasticité aux erreurs. Les deux flux binaires doivent être recombinaés avant décodage.
- 3.35 image D:** type d'image qui ne doit pas être utilisé, sauf dans l'ISO/CEI 11172-2.
- 3.36 coefficient DC:** coefficient DCT pour lequel la fréquence est nulle dans les deux dimensions.
- 3.37 coefficient DCT:** amplitude d'une fonction spécifique à base cosinus.
- 3.38 tampon d'entrée de décodeur:** tampon de type premier entré-premier sorti (FIFO), spécifié dans le vérificateur de mémoire vidéo.
- 3.39 décodeur:** matérialisation d'un processus de décodage.
- 3.40 décodage (processus):** processus défini dans la Rec. UIT-T H.262 | ISO/CEI 13818-2, consistant à lire un flux binaire codé entrant et à produire après décodage des images ou des échantillons audio.
- 3.41 quantification inverse:** processus de remise à l'échelle des coefficients DCT quantifiés, après le décodage de leur représentation dans le flux binaire et avant leur présentation à l'opération de transformation DCT inverse.
- 3.42 support d'enregistrement numérique (DSM) (Digital Storage Media):** dispositif ou système d'enregistrement ou de transmission numérique.

3.43 transformée en cosinus discrète (DCT) (Discrete Cosine Transform): résultat d'une transformation en cosinus, discrète, directe ou inverse. L'opération DCT est une transformation orthogonale discrète, qui peut être inversée. La DCT inverse est définie dans l'Annexe A de la Rec. UIT-T H.262 | ISO/CEI 13818-2.

3.44 format d'affichage: rapport hauteur/largeur (exprimées en unités SI) de l'affichage prévu.

3.45 ordre d'affichage: ordre dans lequel les images décodées sont affichées. Il s'agit normalement du même ordre que celui dans lequel ces images ont été présentées à l'entrée du décodeur.

3.46 processus d'affichage: processus (non normatif) permettant l'affichage des images reconstruites.

3.47 prédiction anticipée à double polarité: mode prédictif consistant à mettre en moyenne deux prédictions monotrames anticipées. La taille de bloc prédite est de 16×16 échantillons de luminance. Les prédictions anticipées à double polarité ne sont utilisées que pour les images à balayage entrelacé.

3.48 édition: opération consistant à manipuler un ou plusieurs flux binaires de données codées afin d'obtenir un nouveau flux binaire de données codées. Les flux binaires convenablement édités doivent toujours répondre aux prescriptions définies dans la Rec. UIT-T H.262 | ISO/CEI 13818-2.

3.49 codeur: matérialisation d'un processus de codage.

3.50 codage (processus de): processus, non spécifié dans la Rec. UIT-T H.262 | ISO/CEI 13818-2, qui effectue la lecture d'un train d'images d'entrée et qui produit un flux binaire valide, tel que défini dans la Rec. UIT-T H.262 | ISO/CEI 13818-2.

3.51 couche d'amélioration: référence relative à une couche (au-dessus de la couche de base) à l'intérieur d'une hiérarchie échelonnée. Indépendamment de la forme d'échelonnabilité, il est possible de définir son processus de décodage par rapport au processus de décodage de la couche inférieure et par rapport au processus de décodage supplémentaire approprié concernant la couche d'amélioration proprement dite.

3.52 lecture rapide en avant: processus d'affichage rapide d'une séquence d'images, ou de parties de séquence d'images, dans leur ordre d'affichage.

3.53 lecture rapide en arrière: processus d'affichage rapide d'une séquence d'images dans l'ordre inverse de leur affichage.

3.54 trame: pour un signal vidéo entrelacé, ensemble de lignes alternées d'une image bi-trame. Une image bi-trame entrelacée se compose donc de deux trames: la trame supérieure et la trame inférieure.

3.55 prédiction monotrame: mode prédictif utilisant seulement une trame de l'image de référence. La taille de bloc prédite correspond à 16×16 échantillons de luminance.

3.56 période de trame: inverse du double de la fréquence d'image.

3.57 image monotrame; image à structure monotrame: image codée dont l'élément structurel `picture_structure` a la valeur «Top field» ou «Bottom field».

3.58 indicateur: variable entière de 1 bit qui peut prendre une valeur parmi deux (zéro ou un).

3.59 interdite: qualificatif d'une valeur qui ne doit jamais être utilisée dans les articles définissant le flux binaire codé. Cela permet généralement d'éviter l'émulation de codes de déclenchement.

3.60 mise à jour forcée: processus par lequel des macroblocs subissent de temps en temps un codage intra pour veiller à ce qu'il n'y ait pas d'accumulation excessive, dans les codeurs et dans les décodeurs, d'erreurs de correspondance entre les processus de transformation DCT inverse.

3.61 compatibilité aval: une nouvelle norme de codage est dite à compatibilité aval avec une ancienne norme de codage si les décodeurs étudiés pour fonctionner avec la nouvelle norme de codage sont en mesure de décoder les flux binaires produits conformément à l'ancienne norme de codage.

3.62 vecteur de mouvement anticipé: vecteur de mouvement qui est utilisé pour la compensation de mouvement à partir d'une ou de deux trames de référence postérieures à ce vecteur dans l'ordre d'affichage.

3.63 prédiction anticipée: prédiction à partir de l'image (trame) de référence passée.

3.64 image bi-trame: ensemble de lignes d'information spatiale acheminées par un signal vidéo. En balayage progressif, ces lignes contiennent des pels qui commencent à partir d'un certain instant et qui forment des lignes successives jusqu'en bas de l'image. En balayage entrelacé, une image bi-trame se compose de deux trames: la trame supérieure et la trame inférieure. L'une de ces trames commencera une période de trame plus tard que l'autre.

- 3.65 prédiction bi-trame:** mode prédictif utilisant les deux trames de l'image de référence.
- 3.66 période d'image:** inverse de la fréquence image.
- 3.67 image bi-trame; image à structure bi-trame:** image codée dont l'élément structurel `picture_structure` a la valeur «Frame».
- 3.68 fréquence image:** fréquence à laquelle les images bi-frames doivent sortir du processus de décodage.
- 3.69 trame de référence postérieure:** trame qui se construit postérieurement à l'image actuelle, dans l'ordre d'affichage.
- 3.70 réordonnement de trame:** processus de réordonnement des images bi-frames reconstruites lorsque l'ordre de codage diffère de l'ordre d'affichage. Il y a réordonnement des images bi-frames en présence d'images à codage bidirectionnel dans un flux binaire. Il n'y a pas de réordonnement des images bi-frames lors du décodage de flux binaires à faible délai.
- 3.71 groupe d'images:** notion définie uniquement dans ISO/CEI 11172-2 (MPEG-1 vidéo). Dans la Rec. UIT-T H.262 | ISO/CEI 13818-2, l'insertion d'en-têtes de groupes d'images permet d'obtenir une fonctionnalité analogue.
- 3.72 en-tête:** bloc de données dans le flux binaire codé, contenant la représentation codée d'un certain nombre d'éléments de données se rapportant aux données codées faisant suite à l'en-tête dans le flux binaire.
- 3.73 échelonnabilité hybride:** combinaison de deux (ou plus de deux) types d'échelonnabilité.
- 3.74 entrelacement:** caractéristique de balayage en télévision conventionnelle où une alternance de lignes de trame représente différents instants sur l'échelle du temps. Dans une image bi-trame entrelacée, l'une des trames est censée être balayée en premier: c'est la première trame, qui peut être la trame supérieure ou la trame inférieure de l'image bi-trame.
- 3.75 image I monotrame:** image I structurée en trame unique.
- 3.76 image I bi-trame:** image I structurée en deux trames.
- 3.77 image I; image à codage intra:** image codée au moyen d'informations ne provenant que de cette image.
- 3.78 codage intra:** codage d'un macrobloc ou d'une image au moyen d'informations ne provenant que de ce macrobloc ou de cette image.
- 3.79 niveau:** ensemble défini de contraintes sur les valeurs que peuvent prendre les paramètres de la Rec. UIT-T H.262 | ISO/CEI 13818-2 dans un profil particulier. Chaque profil peut contenir un ou plusieurs niveaux. Dans un contexte différent, le terme niveau désigne la valeur absolue d'un coefficient non nul (voir «exécution»).
- 3.80 couche:** dans une hiérarchie échelonnée, désigne l'un des ensembles ordonnés de flux binaires et (le résultat de) son processus de décodage (comportant implicitement le décodage de toutes les couches situées au-dessous).
- 3.81 flux binaire de couche:** flux binaire simple associé à une couche spécifique (toujours associé à des qualificatifs de couche, par exemple «flux binaire de couche d'amélioration»).
- 3.82 couche inférieure:** référence relative à la couche située immédiatement au-dessous d'une couche d'amélioration donnée (comportant implicitement le décodage de toutes les couches situées au-dessous de cette couche d'amélioration).
- 3.83 composante de luminance:** matrice, bloc ou pel isolé qui représente les informations monochromes du signal et qui s'applique aux couleurs primaires selon le mode défini dans le flux binaire. Le symbole utilisé pour la luminance (ou clarté) est Y.
- 3.84 Mbit:** symbole du mégabit, soit 1 000 000 bits.
- 3.85 macrobloc:** ensemble des quatre blocs de 8×8 éléments de données de luminance et des deux (en format de chrominance 4:2:0), quatre (en format de chrominance 4:2:2) ou huit (en format de chrominance 4:4:4) blocs de 8×8 éléments de données de chrominance correspondants, issus d'une section de 16×16 éléments de la composante de luminance de l'image. Le terme de *macrobloc* est utilisé pour faire référence parfois aux données échantillonnées et parfois à la représentation codée des valeurs échantillonnées et à d'autres éléments de données définis dans l'en-tête du macrobloc selon la syntaxe indiquée dans la Rec. UIT-T H.262 | ISO/CEI 13818-2. C'est le contexte qui détermine l'usage de ce terme.