

Hasil Penelitian  
Yang Tidak Dipublikasikan

*Image Quality Assessment: Metode Standar dan  
Evaluasinya*



Irwan Prasetya Gunawan  
Guson P. Kuntarto  
Berkah I. Santoso

Fakultas Teknik dan Ilmu Komputer  
Universitas Bakrie  
Jakarta  
2021/2022

## LEMBAR PENGESAHAN HASIL PENELITIAN YANG TIDAK DIPUBLIKASIKAN


1. Judul Penelitian : *Image Quality Assessment*: Metode Standar dan Evaluasinya
2. Peneliti Utama
  - a. Nama Lengkap : Irwan Prasetya Gunawan, Ph.D
  - b. Jenis Kelamin : Laki Laki
  - c. Pangkat/Golongan/NIDN : Lektor / III c / 0301107306
  - d. Bidang Keahlian : Multimedia Processing, Data Communication
  - e. Program Studi : Informatika
3. Tim Peneliti : Guson P. Kuntarto, ST, M.Sc & Berkah I. Santoso, ST, MTI
4. Jangka waktu penelitian : 1 Sept 2021 – 20 Feb 2022

Jakarta, 20 Februari 2022

Menyetujui,

**Ketua Lembaga Penelitian dan  
Pengembangan**

( **Deffi Ayu Puspito Sari, Ph.D.** )  
NIDN: 0308078203

Peneliti Utama  


( **Irwan Prasetya Gunawan, Ph.D.** )  
NIDN: 030110730

# *Image Quality Assessment: Metode Standar dan* Evaluasinya

Irwan Prasetya Gunawan  
Guson P. Kuntarto  
Berkah I. Santoso

## Ringkasan

Pada makalah ini, pengertian kualitas gambar akan dibahas dari hal-hal yang bersifat umum hingga ke metode-metode pengukurannya. Pengaruh subjektivitas manusia sangat terasa dalam penilaian kualitas gambar, sehingga metode pengukuran kualitas secara subjektif juga akan diulas di sini. Metode subjektif ini dapat dianggap sebagai patokan bagi pengukuran kualitas gambar secara umum, namun memiliki kekurangan utama yang menjadikannya kurang praktis untuk bisa dilakukan sewaktu-waktu. Hal inilah yang menjadi alasan utama mengapa metode objektif dikembangkan dan digunakan karena metode yang terakhir ini dapat memberikan hasil yang lebih cepat dan praktis. Secara khusus, makalah ini akan mengulas metode pengukuran kualitas yang ditujukan untuk gambar *Standard Dynamic Range (SDR)*. Selain itu, akan dibahas juga beberapa macam kriteria unjuk kerja metode pengukuran kualitas secara umum.

## 1 Umum

Kualitas gambar adalah hal yang mungkin mudah bagi kita untuk diucapkan, tapi terkadang sulit untuk didefinisikan. Kualitas gambar bisa dilihat dari sudut pandang isi atau konten gambar, tetapi juga bisa dinilai dari kualitas reproduksi atau tampilan visualnya.

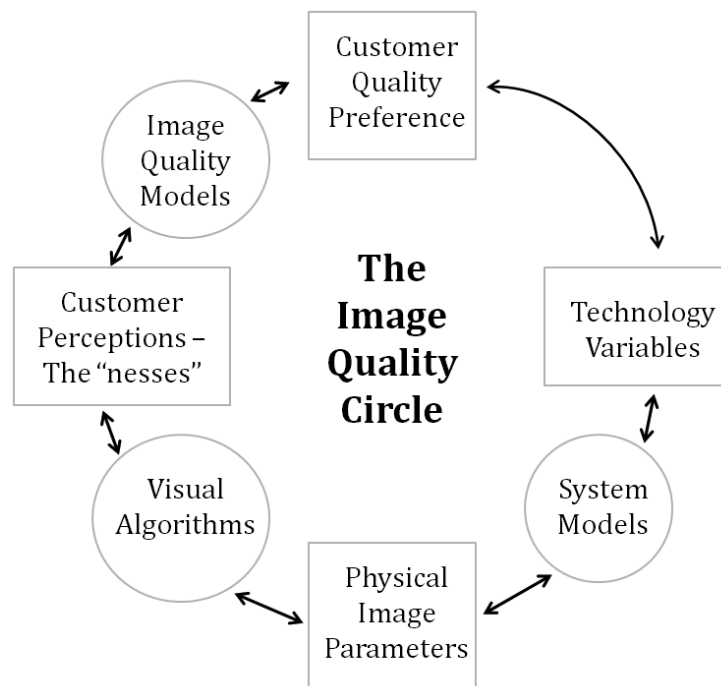
Dalam dunia fotografi, kualitas gambar foto bisa dilihat dari aspek komposisinya. Misalnya, bagaimana sebuah foto bisa membawa kita untuk fokus pada objek tertentu dan bukan pada objek lain yang ada pada foto yang sama. Ini dikenal dengan teknik *framing* dalam fotografi, dan barangkali termasuk salah satu teknik dasar bagi para fotografer yang harus dikuasai untuk bisa menghasilkan sebuah foto atau gambar yang akan dianggap memiliki kualitas yang baik (atau kualitas profesional). Selain itu, komposisi gambar dalam sebuah foto juga dipengaruhi oleh tata letak objek utama dalam bingkai gambar. Dalam hal ini, salah satu aturan klasik tidak tertulis yang sering dianut oleh fotografer adalah

*rule of thirds*, yaitu komposisi yang menempatkan objek utama foto di sepertiga kiri atau kanan gambar dan membiarkan dua pertiga sisanya lebih terbuka.

Komposisi warna dalam sebuah foto juga sering menentukan apakah sebuah foto bisa dianggap sebagai foto yang menunjukkan kualitas profesional atau tidak. Termasuk ke dalam hal ini adalah tata letak warna serta *balance* atau keseimbangan proporsi berbagai warna yang berbeda dalam sebuah gambar. Dalam banyak hal, ini juga terkait dengan kondisi pencahayaan (*lighting*) pada saat gambar foto diambil, yaitu bagaimana arah cahaya datang dalam sebuah frame gambar serta *tone* pencahayaan yang akan menentukan nuansa warna pada foto yang dibuat. Dari sinilah istilah *golden hours* didapatkan, yaitu waktu-waktu tertentu yang bisa memberikan kondisi pencahayaan alami yang baik dan dianggap bisa menghasilkan foto berkualitas. Penelitian yang dilaporkan dalam [1] berusaha untuk mengklasifikasi sebuah foto ke dalam foto profesional (dan dianggap memiliki kualitas yang bagus) atau bukan.

Kualitas visual gambar, di lain pihak, sering dikaitkan dengan bagaimana sebuah gambar direproduksi atau direkonstruksi. Contoh yang mungkin sering kita temukan sehari-hari adalah turunnya kualitas visual gambar ketika gambar tersebut kita kirimkan melalui aplikasi pengirim pesan karena proses kompresi gambar yang dilakukan.

Pengukuran kualitas gambar bisa dilihat sebagai bagian keseluruhan dari konsep Lingkaran Kualitas Gambar (*Image Quality Circle*), menurut [2]. Konsep ini menghubungkan beberapa hal seperti variabel teknologi, preferensi kualitas

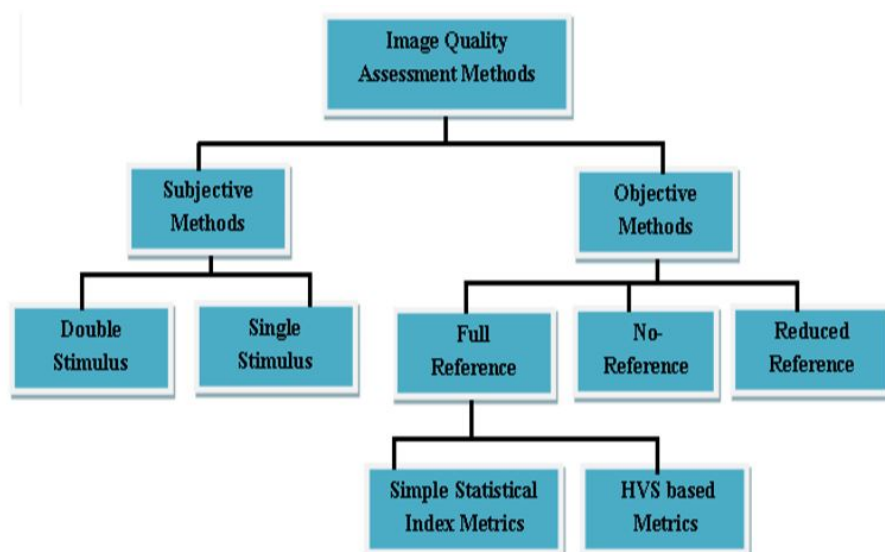


Gambar 1: Lingkaran kualitas gambar; diambil dari [2]

pengguna/pelanggan, serta parameter-parameter fisik gambar (Gambar 1).

Metode pengukuran kualitas gambar, jika dirunut dari sejarahnya, memiliki sejarah yang cukup panjang, namun masih menjadi topik penelitian yang cukup hangat hingga saat ini. Penerapan metode tersebut bisa sangat luas, mulai dari penilaian kualitas teknik pengkodean gambar, pemantauan kualitas layanan, watermarking, peningkatan citra, aplikasi di dunia medis dan hiburan, dan lain-lain. Misalnya, jika sebuah gambar diproses dengan metode kompresi tertentu, maka kita tentu ingin mengevaluasi apakah metode kompresi yang diterapkan berpengaruh terhadap kualitas visual gambar yang dihasilkan. Atau, jika pada gambar ini kita sisipkan pesan rahasia (melalui proses steganografi), maka kita ingin mengetahui apakah proses steganografi yang dilakukan menimbulkan distorsi pada gambar. Adakalanya distorsi pada gambar sebenarnya memang terjadi (dan tidak bisa dihindari), namun mata manusia tidak terlalu sensitif terhadap adanya distorsi ini. Atau, jika gambar yang mengalami distorsi tersebut ditampilkan pada perangkat tampilan tertentu, maka distorsi visual yang terjadi pada gambar mungkin tidak terlalu nyata (dan juga tidak terlalu mengganggu) bagi pengamat atau pengguna.

Selain itu, keterbatasan daya tangkap sistem konvensional (SDR) dimanfaatkan dalam teknologi kompresi data gambar atau video, sehingga informasi visual yang dianggap kurang relevan atau kurang penting bisa diabaikan atau dibuang tanpa harus mengurangi kandungan informasi keseluruhan dengan kualitas visual yang masih bisa diterima oleh pengguna akhir. Ini berarti kita membutuhkan metode pengukuran kualitas yang dapat mewakili tanggapan persepsi kualitas yang diterima oleh pengguna akhir. Umumnya pengukuran kualitas dilakukan dengan metode subjektif yang melibatkan beberapa pengamat. Hal ini tentunya akan mengeluarkan biaya yang tinggi dan sangat tidak



Gambar 2: Taksonomi metode pengukuran kualitas gambar; diambil dari [3, 4]

praktis. Oleh karena itu digunakan metode yang lebih murah, lebih cepat dan mudah untuk diintegrasikan ke suatu sistem ataupun aplikasi, yaitu metode pengukuran kualitas secara objektif.

Taksonomi metode pengukuran kualitas gambar (*Image Quality Assessment (IQA)*) bisa dilihat pada Gambar 2. Penjelasan lebih lanjut akan diberikan pada sub-bagian berikut ini.

## 2 Metode Subjektif

Penilaian kualitas subjektif adalah eksperimen terkontrol dengan partisipan manusia yang bertujuan untuk mengukur persepsi kualitas gambar sesuai dengan persepsi indera penglihatan manusia (*Human Visual Systems (HVS)*). Dalam eksperimen tersebut, *golden standard* untuk tujuan benchmarking adalah penilaian manusia tanpa intervensi dari orang lain [3]. Tidak hanya itu, metode subjektif juga dapat memberikan wawasan tentang perilaku manusia dalam konteks penilaian kualitas gambar [5]. Sudah lama disadari bahwa tugas evaluasi kualitas gambar bagi manusia yang melihatnya tidak hanya melibatkan proses fisiologis tetapi juga ada aspek psikologis dalam proses tersebut. Oleh karena itu, metode subyektif juga cocok untuk digunakan sebagai patokan untuk berbagai algoritma dan metode dalam pemrosesan gambar/video, termasuk algoritma penilaian kualitas gambar.

### 2.1 Cara Investigasi

Beberapa cara untuk menginvestigasi kualitas gambar secara subyektif dan membuat metriknya antara lain dijelaskan dalam [6]:

- Sifat statistik persepsi kualitas: persepsi atau penilaian dapat dilihat sebagai variabel acak dengan distribusi yang dijelaskan oleh semua kemungkinan persepsi atau penilaian individu dalam suatu populasi. Penilaian banyak pengamat manusia yang ditambahkan bersama-sama dapat dianggap sebagai terdistribusi normal.
- *Just Noticeable Differences (JND)*: pada umumnya, untuk membangun metrik kualitas subjektif, formulasinya didasarkan atas perbandingan dua *stimulus* (misalnya, gambar atau bagian gambar yang dibandingkan dengan distorsi yang mungkin muncul di bagian gambar tersebut). *JND* adalah perbedaan stimulus yang mengarah pada proporsi respons positif dan negatif dalam perbandingan berpasangan.
- *Cross modal psychophysics* (psikofisika lintas modal): studi para ahli menunjukkan bahwa ada korelasi antara persepsi manusia dari indera yang berbeda.

- Skala pengukuran subjektif: Skala pengukuran subjektif adalah peringkat kualitas yang diberikan oleh pengamat pada Lingkaran Kualitas Gambar (Image Quality Circle) [2]. Jenis skala mengatur hasil pengukuran subjektif.

Ada banyak metodologi dan aturan yang berbeda untuk merancang tes dalam penilaian kualitas subjektif ini. Ringkasan beberapa penelitian pada topik pengukuran kualitas subjektif bisa dilihat pada Tabel 1.

Tabel 1: Ringkasan beberapa metode penilaian subjektif

Paper	Metode	Deskripsi
Dijk dkk. [7]	Kategori Penskalaan	Teknik penskalaan kategori numerik memberikan efisiensi dan cara yang valid untuk mendapatkan rasio kompresi versus kurva kualitas dan untuk menilai kualitas gambar yang dirasakan dengan cara yang jauh lebih kecil.
NS. Alpert (CCETT) dan J.-P. Evain (EBU) [8]	SSCQE dan DSCQE	SSCQE untuk mengevaluasi kualitas subjektif, sedangkan DCSQE adalah digunakan untuk menjaga kualitas gambar dan informasi ditransmisikan.
Sheikh dkk. [9]	Double Stimulus	Eksperimen menggunakan metodologi stimulus ganda untuk mengukur kualitas lebih akurat untuk tujuan penataan kembali
Redi dkk. [10]	SS dan metode QR	Single Stimulus (SS) memiliki beberapa kelemahan. Metode Quality Ruler (QR) layak untuk diterapkan dari sudut pandang konsistensi dan pengulangan skor
Mantiuk dkk. [11]	Force-choice pairwise comparison	Metode ini menghasilkan varians pengukuran terkecil dan dengan demikian menghasilkan hasil yang paling akurat. Metode ini juga yang paling efisien waktu, dengan asumsi jumlah moderat dibandingkan kondisi.
Persson [6]	QR	Perbedaan penilaian dalam penelitian ini tampaknya signifikan tergantung pada kesamaan yang dirasakan antara pengukuran gambar rujukan dan gambar uji.
Nuutinen et al. [12]	Dynamic Reference	Metode DR sangat cocok untuk eksperimen yang membutuhkan hasil yang sangat akurat dalam waktu singkat karena metode DR lebih akurat daripada metode ACR dan lebih cepat daripada metode PC

*berlanjut ke halaman berikutnya...*

Tabel 1 – lanjutan dari halaman sebelumnya

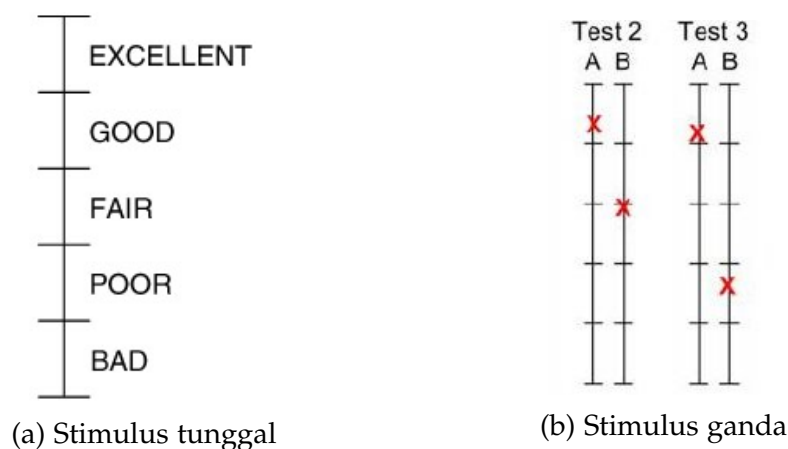
Paper	Metode	Deskripsi
Zhu dkk. [13]	AIT Terinspirasi MOS dan PC	Menggunakan Arrow's Impossibility Theorem (AIT) membuktikan bahwa pertemuan antara kebulatan suara dan kebebasan alternatif yang tidak relevan (IIA) akan menghasilkan "subjek penting", yang pada kenyataannya menentukan peringkat akhir kualitas gambar

Dalam penilaian kualitas gambar secara subjektif, citra ditampilkan kepada sejumlah pengamat yang diminta untuk membandingkan citra asli dengan citra terdistorsi (atau citra pengujian) agar pengamat dapat memberikan skor atau nilai yang mencerminkan kualitas citra yang terdistorsi. Skor dari berbagai responden kemudian dirata-ratakan. Skor rata-rata inilah yang dianggap mewakili kualitas gambar terdistorsi.

## 2.2 Kategori

Metode subyektif dapat dibagi menjadi dua kategori besar [3]: metodologi stimulus tunggal dan metodologi stimulus ganda. Pembagian ini didasarkan atas jumlah gambar yang ditampilkan secara bersamaan kepada pengamat/pemberi nilai. Metode stimulus tunggal, misalnya, menampilkan gambar asli dan gambar uji secara berturut-turut. Ini berbeda dengan metode stimulus ganda yang menampilkan gambar asli dan gambar uji secara bersamaan dalam satu layar yang sama. Dalam beberapa metode, pengamat tidak mengetahui gambar mana yang merupakan gambar asli dan gambar mana yang merupakan gambar uji.

Dalam metode dengan stimulus tunggal seperti *Single Stimulus Continuous Quality Evaluation (SSCQE)*, subjek secara terus-menerus menunjukkan kesan mereka terhadap kualitas video pada skala linier yang dibagi menjadi lima se-



Gambar 3: Skala penilaian subjektif



gmen (Gambar 3a). Metode lainnya yang menggunakan stimulus tunggal antara lain *Absolute Category Rating (ACR)*, *Absolute Category Rating with Hidden Reference removal (ACR-HR)*, dan *Subjective Assessment Methodology for Video Quality (SAMVIQ)* (untuk video). Sementara itu, metode yang menggunakan stimulus ganda seperti *Double Stimulus Continuous Quality Scale (DSCQS)* adalah bentuk metode berbasis diskriminasi dan memiliki keuntungan ekstra bahwa skor subjektif kurang terpengaruh oleh adaptasi dan efek kontekstual. Perbedaan antara skor referensi dan urutan terdistorsi memberikan penilaian penurunan subjektif (Gambar 3b). Metode-metode pengukuran lainnya yang menggunakan stimulus ganda antara lain *Double Stimulus Impairment Scale (DSIS)*, *Simultaneous Double Stimulus for Continuous Evaluation (SDSCE)*, *Degradation Category Rating (DCR)*, atau *Pair Comparison (PC)* [14].

Ada banyak metode subjektif untuk menilai kualitas gambar 2-dimensi statis dan dinamis yang sudah distandardisasikan oleh badan standar dunia seperti ITU-T. Perbandingannya bisa dilihat di Tabel 2. Metode-metode ini banyak yang sudah didokumentasikan dalam berbagai standar dan rekomendasi untuk penilaian kualitas gambar [15]. Standar ini, antara lain, muncul dari kebutuhan dunia industri, termasuk di antaranya adalah industri telekomunikasi. Diterimanya sebuah metode pengukuran kualitas subjektif dalam sebuah standar yang berlaku secara internasional menunjukkan bahwa metode pengukuran subjektif ini telah melalui serangkaian proses yang cukup panjang agar bisa diterima secara luas.

Tabel 2: Perbandingan metode pengukuran kualitas subjektif; diambil dari [14]

Metode	Skor	Stimulus	Simultan	Pengulangan	Gambar rujukan	Cocok untuk banyak pengamat
DSIS	5-grade impairment	Double	No	Optional	Yes	Yes
DSCQS	Continuous	Double	No	Optional	Yes	Optional
SS	5-grade quality, 11-grade quality, non-category judgement	Single	No	No	Optional	Yes
SSMR	5-grade quality, 11-grade quality, non-category judgement	Single	No	Yes	Hidden	Yes
SC	7-grade comparison, non-category judgement	Double	Yes	Optional	Yes	Yes
SSCQE	Continuous	Single	No	Optional	No	Yes
SDSCE	Continuous	Double	Yes	Optional	Yes	Yes

*berlanjut ke halaman berikutnya...*

Tabel 2 – lanjutan dari halaman sebelumnya

Metode	Skor	Stimulus	Simultan	Pengu- langan	Gambar rujukan	Cocok untuk banyak penga- mat
SAMVIQ	Continuous	Single	No	Yes	Yes, hi- dden	No
ACR	5-grade quality	Single	No	No	Optional	Yes
ACR- HR	5-grade quality	Single	No	No	Hidden	Yes
DCR	5-grade impair- ment	Double	No	Optional	Yes	Yes
DCR-SP	5-grade impair- ment	Double	Yes	Optional	Yes	Yes
PC	7-grade compari- son, non-category judgement	Double	No	Optional	Yes	Yes
PC-SP	7-grade compari- son, non-category judgement	Double	Yes	Optional	Yes	Yes

### 2.3 Skor Penilaian

Skor penilaian metode pengukuran kualitas subjektif bisa dinyatakan dalam beberapa jenis, antara lain:

- *Mean Opinion Score (MOS)*: rata-rata aritmetika skor numerik individu dari kualitas gambar yang dirasakan oleh pengamat [16]. Biasanya dinyatakan sebagai angka tunggal antara 1 dan 5 yang masing-masing sesuai dengan kualitas yang dirasakan mulai dari nilai yang terendah ( $MOS = 1$ ) hingga ke yang tertinggi ( $MOS = 5$ ). Skor tersebut mewakili peringkat kualitas gambar yang terdegradasi dibandingkan dengan gambar rujukan (yang dapat ditampilkan kepada pemirsa sebelum, sesudah, atau pada saat yang sama/berdampingan dengan video yang terdegradasi dalam tampilan video).
- *Differential Mean Opinion Score (DMOS)*: rata-rata aritmetika dari perbedaan antara skor yang diberikan pada gambar yang terdegradasi dan skor yang diberikan pada gambar rujukan.
- *Mean Subjective Rank (MSR)*: nilai peringkat rata-rata dari eksperimen subjektif yang menampilkan gambar asli dan gambar terdistorsi dengan pengamat yang diminta untuk memberikan peringkatnya menurut kualitasnya [17]. Semakin kecil nilai MSR, semakin baik kualitas gambarnya.

## 2.4 Kelemahan

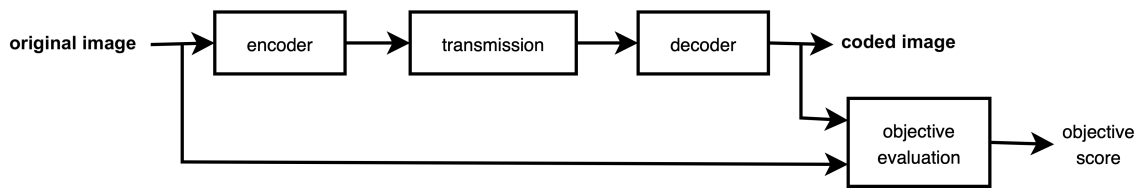
Metode subjektif bukan tanpa kekurangannya sendiri. Ada beberapa masalah yang terkait dengan metode subjektif [18]. Pertama-tama, metode ini mungkin membutuhkan waktu lebih lama untuk diproses, belum lagi juga mahal. Karena eksperimen subjektif diatur untuk membuat subjek mengevaluasi setiap gambar dalam kumpulan data, mungkin diperlukan waktu berjam-jam untuk menyelesaikannya. Untuk mencapai validitas statistik guna keperluan evaluasi, jumlah pemirsa manusia yang terlibat dalam eksperimen juga harus cukup besar sehingga hasil tidak diperoleh secara kebetulan (*chance*). Namun, akhir-akhir ini, metode *crowdsourcing* juga telah digunakan untuk melibatkan lebih banyak pemirsa dalam proses evaluasi dalam waktu yang jauh lebih singkat daripada dalam eksperimen subjektif tradisional [19]. Metode berbasis *crowdsourcing* tersebut bukannya tanpa tantangan; misalnya, tidak seperti eksperimen subjektif tradisional di lingkungan laboratorium, hanya ada kontrol terbatas atau bahkan tidak ada pengaturan eksperimental (perangkat tampilan, kondisi pencahayaan, jarak pandang, dll).

Terlepas dari pengaturan eksperimental yang digunakan, evaluasi subjektif mahal karena pengamat sebagai subjek uji juga biasanya harus direkrut dan dibayar. Eksperimen subjektif tradisional mungkin lebih mahal karena pengukurannya mungkin memerlukan pengaturan laboratorium yang mungkin sulit diatur dengan peralatan khusus yang terkalibrasi. Evaluasi subjektif mungkin juga tidak cocok untuk aplikasi tertentu [20]; misalnya, situasi *real time* yang membutuhkan adanya tanggapan langsung.

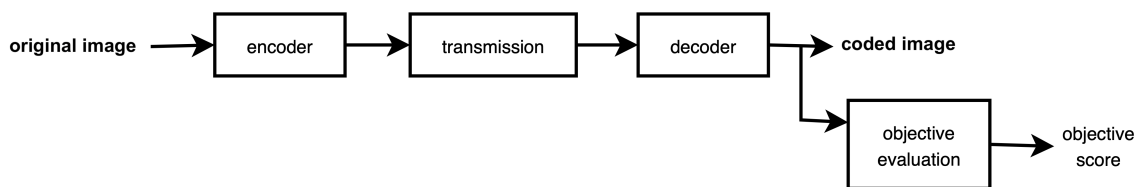
Permasalahan tersebut menjadi alasan utama mengapa banyak peneliti beralih ke metode objektif yang dapat memberikan hasil yang lebih cepat dan praktis [5, 21].

## 3 Metode Objektif

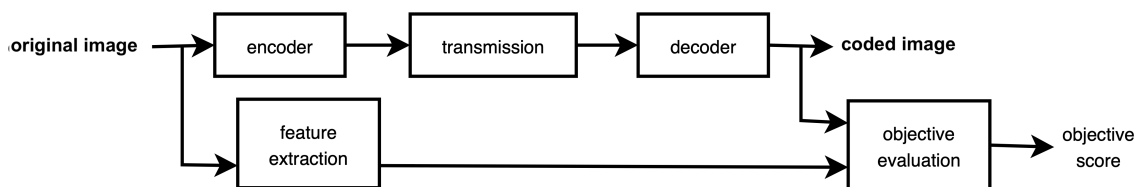
Pengukuran kualitas visual secara objektif ini bisa dilakukan dengan berbagai metode komputasi yang memanfaatkan faktor-faktor HVS, kontras, warna, struktur, error/distorsi pada gambar, dan sebagainya. Beberapa jenis distorsi yang mungkin timbul pada gambar untuk pengukuran kualitas ini antara lain distorsi buram (*blur*), buram karena pergerakan, tepian, kontur, artefak blok, *granular noise*, gerakan yang terpatah-patah, dan sebagainya. Evaluasi dinyatakan sebagai rumus matematika yang dapat dihitung tanpa campur tangan manusia. Untuk mendapatkan evaluasi yang lebih baik, skor subjektif dari eksperimen subjektif dapat digunakan sebagai acuan untuk model objektif ini. Kemudahan yang didapat dalam melakukan proses komputasi ini (apabila dibandingkan dengan metode subjektif) membuat pengukuran objektif semakin populer. Sifat praktis metode penilaian kualitas gambar objektif menjadikannya sangat cocok untuk berbagai aplikasi seperti sistem kontrol kualitas, *benchmark* algoritma pemrosesan gambar, dan optimasi sistem transmisi.



(a) Model dengan rujukan penuh, **FR**



(b) Model tanpa rujukan, **NR**

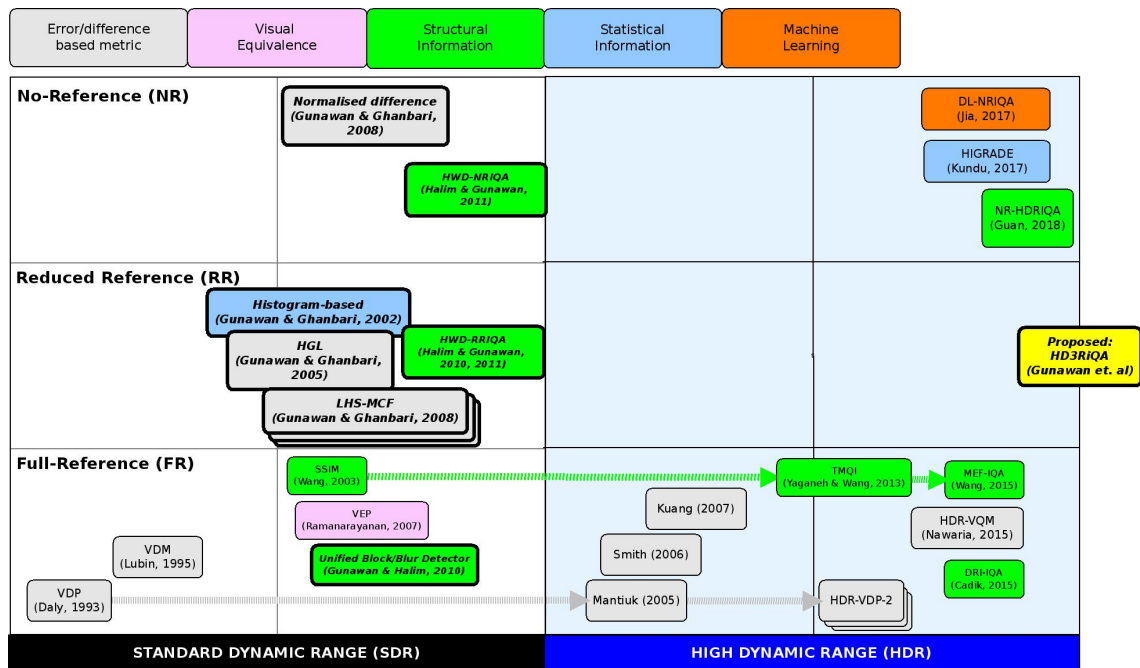


(c) Model dengan rujukan sebagian/tereduksi, **RR**

Gambar 4: Model pengukuran kualitas gambar secara objektif

Seperti yang ditunjukkan pada Gambar 2, teknik pengukuran kualitas gambar digital dengan metode objektif dapat dibagi menjadi tiga pendekatan utama. Menurut *Video Quality Expert Group (VQEG)*, tiga kategori dasar untuk metode penilaian kualitas gambar/video didasarkan pada ketersediaan gambar referensi yang digunakan sebagai rujukan dalam penilaian. Kategori yang dimaksud adalah referensi penuh (*full-reference, Full-Reference (FR)*), referensi tereduksi (*reduced-reference, Reduced-Reference (RR)*), dan metode tanpa referensi (*no-reference, No-reference (NR)*) [16, 22, 23]. Ilustrasinya bisa dilihat pada Gambar 4.

Metode referensi penuh (**FR**) didefinisikan oleh [24] sebagai metode di mana pengukuran dilakukan pada perbedaan antara referensi dan gambar kode. Perbedaannya dapat digunakan untuk menghitung ukuran distorsi komparatif. Ini ditunjukkan pada Gambar 4a. Metode ini mengevaluasi kualitas gambar dengan cara membandingkan gambar uji dengan versi asli yang tidak terdistorsi [14]. Model kualitas gambar No-Reference (**NR**), di sisi lain, mencoba untuk meniru bagaimana **HVS** atau mata manusia melihat kualitas gambar tanpa memerlukan gambar referensi asli. Metode ini sering juga disebut sebagai penilaian kualitas gambar buta [3]. Ini direpresentasikan oleh Gambar 4b. Selanjutnya, penilaian kualitas gambar dengan metode *reduced-reference (RR)* memberikan solusi yang seimbang dan *trade-off* antara dua ekstrem yang diwakili oleh model kualitas **FR** dan **NR**. Metode **RR** dirancang untuk hanya menggunakan sebagian data ten-



Gambar 5: Peta jalan kategori pengukuran kualitas gambar

tang gambar referensi untuk mengevaluasi gambar yang diproses. Data parsial dapat dibentuk dari fitur yang diekstraksi dari sinyal yang tidak terdistorsi yang kemudian dibandingkan dengan fitur yang diekstraksi dari gambar yang diproses atau mengalami degradasi [25]. Penilaian kualitas RR awalnya diusulkan untuk melacak perubahan kualitas visual yang mungkin ada dalam informasi gambar (atau video) yang didistribusikan melalui jaringan komunikasi.

Sebagai metode yang menggunakan data overhead untuk tujuannya, evaluasi kualitas RR berkaitan dengan kecepatan data yang digunakan untuk mengirimkan informasi sampingan ini. Jika, misalnya, saluran tambahan (*side channel*) dengan kecepatan data tinggi ini tersedia, maka metode RR dapat menggunakan jumlah informasi yang lebih besar tentang gambar referensi. Jika *side channel* cukup besar, kita bahkan dapat mengirimkan seluruh gambar referensi asli. Di sisi lain, jika laju data *side channel* ini kecil, metode RR harus dapat bekerja dengan menggunakan informasi sampingan dalam jumlah yang sedikit/terbatas.

Selain pengkategorian berdasarkan ketersediaan gambar rujukan, metode pengukuran kualitas citra objektif dapat juga dibedakan berdasarkan teknik serta metode kuantifikasi yang digunakan untuk mengukur kualitas citra. Kuantifikasi dapat didasarkan pada kesalahan perbedaan [26], informasi struktural [5, 27, 28], dan juga *machine learning* [29].

Metrik kualitas berdasarkan perbedaan kesalahan [26] dapat memanfaatkan berbagai metode/ algoritma pemrosesan gambar baik dalam domain spasial maupun frekuensi. Kualitasnya kemudian dikuantifikasi berdasarkan analisis domain spasial-temporal atau frekuensi ini. Metode lain seperti [5, 27] berdasarkan informasi kesamaan struktural menggunakan analisis multi-skala untuk meng-

ukur kualitas sinyal. Metode ini menggunakan metrik *structural similarity index* (SSIM) yang dimodifikasi dengan pendekatan *natural scene statistic* (NSS). Kemudian baru-baru ini, juga ada pendekatan seperti [29] menggunakan *machine learning* berbasis peta *saliency* untuk meningkatkan performa metode NR. Model seperti itu bukannya tanpa masalah; misalnya, adanya kesenjangan yang signifikan dalam nilai luminansi ketika model seperti itu diterapkan pada gambar HDR.

Berbagai macam penelitian yang berfokus pada metode-metode pengukuran kualitas ini digambarkan secara ringkas dalam peta jalan kategori pengukuran kualitas gambar (Gambar 5). Metode pengukuran kualitas visual gambar/video yang rata-rata dikembangkan sebelum tahun 2010 pada umumnya ditujukan untuk pengukuran kualitas visual pada jangkauan teknologi SDR.

## 4 Metrik Kualitas Gambar SDR

Keberanekaragaman metode penilaian kualitas seperti yang sudah dijelaskan di bagian sebelumnya terjadi karena tiap-tiap metode tersebut dapat diterapkan pada beragam skenario aplikasi yang berbeda-beda. Di antara metode objektif tersebut, metrik yang banyak digunakan untuk citra digital adalah *Peak Signal-to-Noise Ratio* (PSNR) (rasio sinyal terhadap noise puncak) atau *Mean Squared Error* (MSE) (rata-rata dari kuadrat selisih).

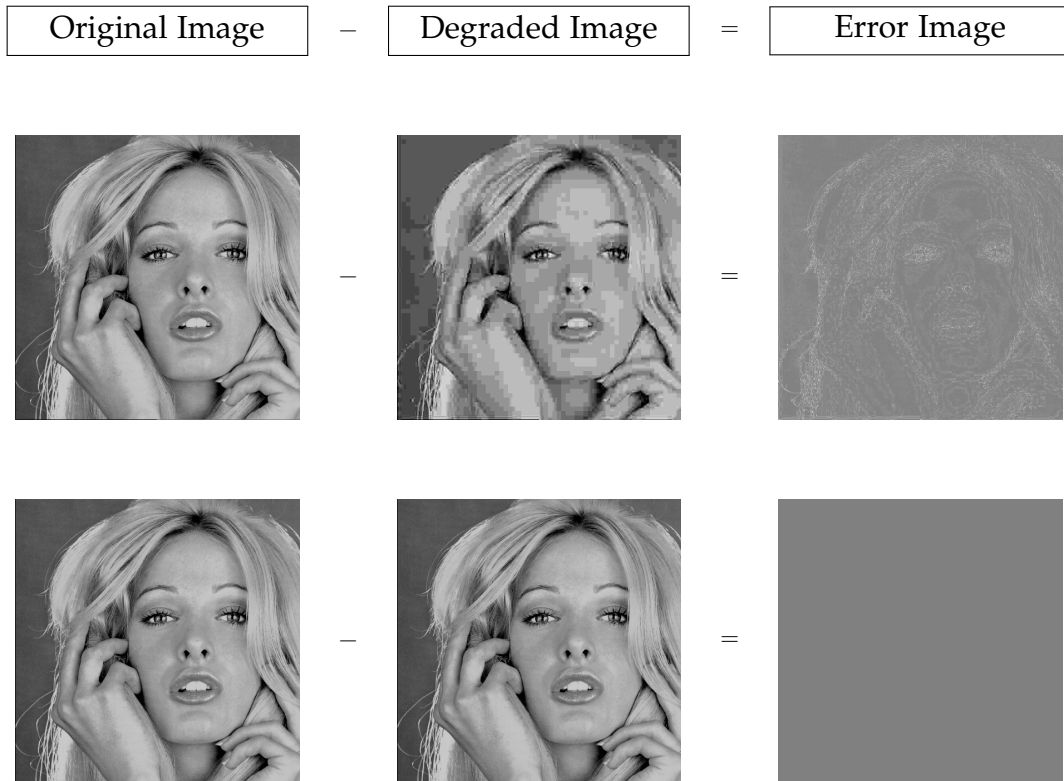
Perbandingan berbasis piksel seperti MSE/PSNR biasanya digunakan dalam skenario evaluasi menggunakan rujukan lengkap. Perbandingan dalam pengukuran referensi penuh dilakukan dengan menghitung sinyal kesalahan, dihitung berdasarkan sampel demi sampel dan bingkai demi bingkai, seperti yang diilustrasikan dalam Gambar 6. Yang harus diperhatikan dalam perhitungan ini adalah sinyal kesalahan harus dihitung setelah sinkronisasi antara gambar rujukan dan gambar uji dilakukan; artinya, kedua gambar harus memiliki dimensi yang sama, serta berisi konten atau objek yang juga sama. Nilai selisih mentah atau selisih mutlak dapat digunakan sebagai indikator sinyal kesalahan, yang amplituda dan sifat-sifat statistiknya dapat digunakan untuk mendapatkan informasi yang representatif mengenai distorsi yang terjadi pada gambar uji.

MSE sebuah gambar (uji) dengan lebar  $N$  piksel dan tinggi  $M$  piksel didefinisikan sebagai

$$\text{MSE} = \frac{\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \left( I^{(r)}(x, y) - I^{(d)}(x, y) \right)^2}{NM} \quad (1)$$

Di sini,  $I^{(r)}(x, y)$  dan  $I^{(d)}(x, y)$  masing-masing adalah nilai luminansi piksel gambar asli (rujukan) dan piksel gambar uji. Variasi dari MSE dengan mengam-





Gambar 6: Menghitung selisih dua buah gambar: dengan perbedaan kualitas yang nyata (atas), dan perbedaan kualitas yang sedikit (bawah).

bil logaritma dari nilainya dikenal sebagai **PSNR**, dan didefinisikan sebagai

$$\text{PSNR(dB)} = -10 \log \left( \frac{\text{MSE}}{\max\{I(x,y)\}^2} \right) = -10 \log \left( \frac{\text{MSE}}{255^2} \right) \quad (2)$$

dengan  $\max\{\dots\}$  adalah fungsi yang mengembalikan nilai tertinggi (maksimum) nilai argumen inputnya. Diasumsikan bahwa nilai maksimum  $I(x,y)$  adalah  $2^8 - 1 = 255$  karena kedalaman piksel adalah 8 bit per piksel. Formulasi beberapa metrik kualitas yang diturunkan dari **MSE** ini bisa dilihat pada Tabel 3.

Tabel 3: Metode pengukuran kualitas gambar berbasis MSE dan PSNR

No.	Metrik	Perhitungan
1	Mean Square Error ( <i>MSE</i> )	$\text{MSE} = \frac{1}{NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} (I^{(r)}(x,y) - I^{(d)}(x,y))^2 \quad (3)$
2	Root Mean Square ( <i>RMS</i> )	$\text{RMS} = \sqrt{\text{MSE}} \quad (4)$

*berlanjut ke halaman berikutnya...*

Tabel 3 – lanjutan dari halaman sebelumnya

No.	Metrik	Perhitungan
3	Normalised Mean Square Error (NMSE)	$NMSE = \frac{\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} (I^{(r)}(x, y) - I^{(d)}(x, y))^2}{\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} (I^{(r)}(x, y))^2} \quad (5)$
4	Signal to Noise Ratio (SNR)	$SNR(\text{dB}) = -10 \log(NMSE) \quad (6)$
5	Peak Signal to Noise Ratio (PSNR)	$PSNR(\text{dB}) = -10 \log \left( \frac{MSE}{\max\{I(x, y)\}^2} \right) = -10 \log \left( \frac{MSE}{255^2} \right) \quad (7)$
6	Signal to Error Ratio (SER)	$SER_{\text{field}} = 20 \log_{10} \left\{ \frac{\max\{I^{(r)}(x, y)\}}{RMS_{\text{space}}(I^{(r)}(x, y) - I^{(d)}(x, y))} \right\} \quad (8)$ $SER_{\text{sequence}} = 20 \log_{10} \left\{ \frac{\max\{I^{(r)}(x, y)\}}{RMS_{\text{time}}(RMS_{\text{space}}(I^{(r)}(x, y) - I^{(d)}(x, y)))} \right\} \quad (9)$
7	Standard Deviation (STD)	$\sigma_{\text{field}}^2 = \frac{1}{NM} \sum_{x=1}^N \sum_{y=1}^M ((I_t^{(r)}(x, y) - I_t^{(d)}(x, y)) - \mu)^2 \quad (10)$ <p>dengan</p> $\mu = \frac{1}{NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} (I^{(r)}(x, y) - I^{(d)}(x, y))$

Metrik kualitas objektif pada Tabel 3 menggunakan asumsi bahwa nilai kesalahan yang besar akan berkontribusi pada kualitas yang lebih rendah. Selain itu, metrik-metrik di atas berlaku untuk gambar dengan *error* yang sifatnya aditif dan tidak memiliki korelasi dengan konten gambar. Masalah akan timbul ketika kita memiliki *error* dengan jumlah yang sedikit namun memiliki amplitudo yang tinggi, atau sebaliknya, yaitu ketika ada banyak *error* dengan amplitudo yang rendah sehingga *error* ini mungkin tidak akan terlihat secara subjektif. Tambahkan lagi, Winkler [30] berpendapat bahwa parameter ini mengukur *fidelity* bukan kualitas karena metrik-metrik di atas menghitung seberapa mirip gambar uji dengan gambar aslinya. Selain itu, MSE dan PSNR sendiri kurang mencerminkan persepsi manusia karena korelasinya dengan kualitas subjektif sangat rendah. Contohnya diilustrasikan pada Gambar 7 yang memperlihatkan beberapa gambar dengan kualitas subjektif yang berbeda, namun memiliki nilai PSNR yang sama.

MSE dan PSNR adalah ukuran kualitas gambar yang umumnya banyak digunakan karena metode perhitungannya yang sangat sederhana. Namun, kedua ukuran tersebut memerlukan suatu gambar referensi (termasuk ke dalam metode FR) yang dalam hal praktis sering kali tidak dapat tersedia. Turunan dari metode MSE/PSNR ini banyak dan rata-rata memanfaatkan tingkat *visibilitas*





(a) Gambar dengan kualitas subjektif buruk. PSNR = 25.11



(b) Gambar dengan kualitas subjektif sedang. PSNR = 25.11



(c) Gambar dengan kualitas baik. PSNR = 25.12

Gambar 7: Gambar dengan kualitas subjektif yang berbeda, namun memiliki nilai PSNR yang sama. Gambar diambil dari [31].

error pada gambar jika dilihat oleh mata manusia. Metode pengukuran seperti VDP (*Visual Differences Predictor*) [32] dan VDM (*Video Distortion Meter*) [33] adalah turunan dari metode berbasis visibilitas error ini.

Sebagai alternatif terhadap metode berbasis visibilitas error ini, ada pula pendekatan pengukuran kualitas gambar dengan metode FR berdasarkan ukuran distorsi struktural; misalnya UQI (*Universal Quality Index*) [17] dan SSIM (*Structural Similarity*) [34]. Kedua teknik tersebut cukup banyak digunakan karena memiliki hasil yang baik, tetapi terbatas untuk gambar yang memiliki referensi.

Untuk mengukur kualitas gambar secara objektif menggunakan pendekatan No-Reference (NR), pengukuran distorsi yang terjadi pada gambar adalah salah satu dari banyak teknik yang digunakan. Bentuk distorsi utama yang dapat terlihat pada gambar adalah *blur* dan *blockiness*. Sebuah teknik dengan pendekatan NR untuk mengukur *blur* secara perceptual dapat dilakukan dengan mengukur kelebaran tepian/*edge* [35]. Di lain pihak, teknik untuk mengukur *blockiness* yang terjadi pada gambar dapat dilakukan berdasarkan karakteristik lokal dari gambar [36]

Di antara kedua ekstrem FR dan NR, teknik pengukuran kualitas gambar lainnya adalah dengan pendekatan RR, misalnya berdasarkan fitur yang diekstraksi dari area lokal pada gambar [37, 38].

## 5 Evaluasi Kinerja Model Kualitas Gambar Objektif

Evaluasi kinerja model kualitas objektif biasanya dilakukan dengan mengikuti metode yang direkomendasikan oleh VQEG [16]. Ada dua kinerja model yang banyak digunakan: *Pearson correlation* dan *Spearman Correlation*. Korelasi Pearson digunakan untuk mengevaluasi keakuratan model. Untuk dua set data, skor subjektif dan output dari model objektif, korelasi Pearson menunjukkan hubungan linier di antara keduanya. Ini dihitung dengan memasang dua set data dan menghitung koefisien momen produk mereka. Kriteria kinerja lain yang banyak digunakan untuk mengevaluasi model objektif adalah korelasi urutan peringkat Spearman (atau singkatnya korelasi Spearman). Korelasi Spearman digunakan untuk mengevaluasi kemonotonan prediksi dari model objektif yang diberikan dan untuk menunjukkan sejauh mana kesepakatan antara skor subjektif dan model objektif dalam hal tanda perubahan kualitas gambar. Untuk setiap datum dalam dua set data, diberikan nomor sesuai dengan peringkat setiap datum dalam set datanya sendiri. Dari sini, perbedaan peringkat statistik datum yang sesuai dapat dihitung. Tingkat persetujuan yang tinggi yang dinyatakan dalam korelasi Spearman menunjukkan bahwa perbedaan dalam peringkat statistik adalah kecil; yaitu, urutan sejumlah besar data dalam dua kumpulan data adalah sama.

Hasil korelasi dapat berkisar dari  $-1$  hingga  $1$ . Nilai  $1$  menunjukkan hubungan linier antara dua kumpulan data. Yang harus diketahui untuk nilai korelasi ini adalah bahwa tanda negatif dari nilai korelasi ini tidak ada hubungannya

dengan model yang tidak sesuai dengan kualitas subjektif/visual; tanda korelasi menyiratkan “arah” asosiasi, yaitu, korelasi negatif berarti bahwa skor yang relatif tinggi pada satu variabel dipasangkan dengan skor yang relatif rendah pada variabel lainnya. Di sisi lain, korelasi positif berarti bahwa skor yang relatif tinggi pada satu variabel dipasangkan dengan skor yang relatif tinggi pada variabel lain, dan skor rendah dipasangkan dengan skor yang relatif rendah. Kesesuaian yang lemah antara kedua dataset ditunjukkan oleh nilai korelasi yang mendekati nol.

Selain korelasi yang disebutkan di atas, ada juga beberapa metrik lain yang dapat digunakan untuk menilai kinerja model objektif dalam hal *prediksi akurasi*, *prediksi monotonisitas*, dan *prediksi konsistensi*. Definisi metrik-metrik tersebut mengikuti rekomendasi dari VQEG [16].

Metrik tambahan ini biasanya bergantung pada rentang dinamis yang sama dari keluaran model dan data subjektif, yang mungkin tidak berlaku untuk beberapa model objektif. Dalam kebanyakan kasus, kriteria kinerja dalam hal hubungan antara data objektif dan subjektif (seperti yang diberikan oleh korelasi Pearson dan Spearman) sudah cukup.

## 5.1 Prediksi Akurasi

*Prediksi akurasi* dari model objektif adalah kemampuan model untuk memprediksi peringkat kualitas subjektif dengan kesalahan rendah. Mengikuti VQEG [16], prediksi akurasi diberikan oleh korelasi linier Pearson antara  $DMOS_p$  (prediksi DMOS) dan DMOS (peringkat kualitas subjektif). Perhitungannya diberikan oleh

$$r_p = \frac{\text{covar}(DMOS, DMOS_p)}{\sigma_{DMOS} \cdot \sigma_{DMOS_p}} \quad (11)$$

dengan

covar : Kovarian antara dua kumpulan data

$\sigma$  : Standar deviasi

DMOS : peringkat kualitas subjektif

$DMOS_p$  : prediksi peringkat kualitas subjektif

Kovarians untuk dua set data dengan jumlah data  $N$  dapat dihitung menurut

$$\text{covar}(DMOS, DMOS_p) = \frac{1}{N} \sum_{i=1}^N (DMOS[i] - \overline{DMOS})(DMOS_p[i] - \overline{DMOS_p}) \quad (12)$$

dan standar deviasi dari setiap dataset dihitung sebagai

$$\sigma_{DMOS} = \sqrt{\frac{1}{N} \sum_{i=1}^N (DMOS[i] - \overline{DMOS})^2} \quad (13)$$

dan

$$\sigma_{\text{DMOS}_p} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{DMOS}_p[i] - \overline{\text{DMOS}_p})^2} \quad (14)$$

dengan

$$\overline{\text{DMOS}} = \frac{1}{N} \sum_{i=1}^N \text{DMOS}[i] \quad (15)$$

$$\overline{\text{DMOS}_p} = \frac{1}{N} \sum_{i=1}^N \text{DMOS}_p[i] \quad (16)$$

adalah rata-rata dari data subyektif dan data yang diprediksi. Korelasi dihitung setelah keluaran model objektif ditransformasikan ke satu set prediksi DMOS oleh fungsi pemetaan non-linier agar sesuai dengan peringkat kualitas subyektif. Salah satu dari fungsi polinomial dengan derajat-3 yang diberikan oleh

$$\text{DMOS}_p = b_0 + b_1 \cdot q_0 + b_2 \cdot (q_0)^2 + b_3 \cdot (q_0)^3 \quad (17)$$

atau fungsi logistik dengan empat parameter seperti

$$\text{DMOS}_p = \frac{b_1 - b_2}{1 + \exp\left\{\frac{b_3 - q_0}{|b_4|}\right\}} + b_2 \quad (18)$$

dapat digunakan sebagai fungsi pemetaan non-linier. Dalam Persamaan (17) dan (18),  $q_0$  adalah keluaran 'mentah' dari model objektif.

Kita juga dapat menghitung korelasi antara keluaran model (tanpa pemetaan non-linier) dan peringkat subyektif, dan membandingkannya dengan analisis non-linier di atas untuk melihat apakah korelasi yang lebih tinggi terhadap data subyektif dapat diperoleh melalui pemetaan non-linier. Oleh karena itu, kecocokan linier sederhana juga dapat digunakan untuk data objektif; yaitu

$$\text{DMOS}_p = a_0 + a_1 \cdot q_0 \quad (19)$$

Kecocokan non-linier dari (17) dan (18) hanya akan digunakan jika kecocokan linier gagal, yang biasanya ditunjukkan oleh perbedaan yang signifikan antara hasil korelasi Pearson dan Spearman.

## 5.2 Prediksi Monotonisitas

*Prediksi monotonisitas* menggambarkan sejauh mana prediksi model sesuai dengan besaran relatif dari peringkat kualitas subyektif. Kriteria ini mengukur jika kenaikan (penurunan) dalam satu kumpulan data dikaitkan dengan peningkatan (penurunan) di kumpulan data lainnya, terlepas dari besarnya kenaikan (penurunan) [lihat 20].

Koefisien korelasi spearman didefinisikan sebagai

$$r_s = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)} \quad (20)$$

dengan  $d_i$  adalah perbedaan peringkat untuk pasangan  $i^{\text{th}}$  dari DMOS dan DMOS<sub>p</sub> yang sesuai dan  $N$  adalah jumlah titik di setiap kumpulan data.

### 5.3 Konsistensi Prediksi

Untuk menilai konsistensi prediksi dari model objektif, kita dapat menggunakan rasio jumlah titik outlier dibandingkan dengan total jumlah titik yang ada. Titik outlier adalah titik yang memenuhi

$$\left| \text{DMOS}[i] - \text{DMOS}_p[i] \right| > 2\sigma_{\text{DMOS}} \quad (21)$$

Rasio outlier dihitung sebagai

$$\text{Rasio Outlier} = \frac{\text{jumlah total keseluruhan outlier}}{N} \times 100\% \quad (22)$$

### 5.4 Kesepakatan Prediksi

Metrik yang berkaitan dengan kesepakatan adalah koefisien  $\kappa$ , dihitung pada data terkuantisasi. Koefisien  $\kappa$  diberikan oleh

$$\kappa = \frac{\sum_{c=1}^5 f_o(c) - \sum_{c=1}^5 f_E(c)}{N - \sum_{c=1}^5 f_E(c)} \quad (23)$$

dengan  $f_o$  adalah jumlah kesepakatan yang diamati antara DMOS dan DMOS<sub>p</sub> untuk masing-masing dari 5 kelas MOS, dan  $f_E$  adalah jumlah kesepakatan karena kebetulan, dan  $N$  adalah jumlah total pengamatan. Nilai  $\kappa$  berada di antara

Tabel 4: Kekuatan kesepakatan berdasarkan nilai  $\kappa$ ; diambil dari [39].

$\kappa$	Penafsiran
$< 0$	<i>poor</i> (buruk)
0 – 0.2	<i>slight</i> (sedikit)
0.21 – 0.4	<i>fair</i> (cukup)
0.41 – 0.6	<i>moderate</i> (sedang)
0.61 – 0.8	<i>substantial</i> (besar)
0.81 – 1.0	<i>almost perfect</i> (hampir sempurna)

Tabel 5: Histogram dua dimensi dari distribusi DMOS dan DMOS<sub>p</sub>, diklasifikasikan ke dalam lima kelas

	DMOS 1	DMOS 2	DMOS 3	DMOS 4	DMOS 5	Total
DMOS <sub>p</sub> 1	$f_o(1)$					$T_p 1$
DMOS <sub>p</sub> 2		$f_o(2)$				$T_p 2$
DMOS <sub>p</sub> 3			$f_o(3)$			$T_p 3$
DMOS <sub>p</sub> 4				$f_o(4)$		$T_p 4$
DMOS <sub>p</sub> 5					$f_o(5)$	$T_p 5$
Total	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$N$

−1 dan 1, tetapi tidak harus ditafsirkan sebagai koefisien korelasi. Nilai sekitar 0,4 menunjukkan bahwa metode ini efisien [16]. Kekuatan kesesuaian antara keluaran model objektif dan data subjektif dapat ditafsirkan seperti disajikan pada Tabel 4 [39].

Nilai  $\kappa$  dihitung setelah nilai DMOS dan DMOS<sub>p</sub> diklasifikasikan ke dalam sejumlah kelas (yaitu, dengan mengkuantisasi skor). Misalnya, 5 kelas dapat digunakan untuk mengklasifikasikan skor seperti diilustrasikan pada Tabel 5). Jumlah observasi di mana DMOS dan DMOS<sub>p</sub> sesuai diberikan di diagonal utama tabel. Bagian dari kesepakatan ini bisa saja dihasilkan secara kebetulan;  $\kappa$  mengoreksi ini dengan menghapus pengamatan yang disebabkan oleh kebetulan (ditunjukkan dengan pengurangan  $f_E$  dalam Persamaan (23)). Dari Tabel 5,  $f_E(c)$  dihitung sebagai hasil kali observasi pada setiap kelas, baik dari data subjektif (kolom) dan objektif (baris); yaitu

$$f_E(c) = \frac{T_c \times T_{pc}}{N}. \quad (24)$$

Interpretasi dari  $\kappa$  terkadang bisa menjadi rumit karena  $\kappa$  mungkin rendah meskipun ada tingkat kesepakatan yang tinggi. Hal ini terjadi karena  $\kappa$  bergantung pada asumsi yang diambil untuk prosedur pengambilan keputusan, misalnya, metode kuantisasi yang digunakan dalam klasifikasi skor. Sebagai contoh, dalam mengkuantisasi skor DMOS<sub>p</sub> ke dalam kelas yang berbeda (kolom pertama pada Tabel 5), kita harus memperhitungkan karakteristik non-linier dari metrik tujuan selain memilih langkah kuantisasi yang sesuai. Pengaturan langkah kuantisasi yang berbeda dapat menghasilkan nilai  $\kappa$  yang berbeda pula. Tambahan lagi, perhitungan  $\kappa$  juga mengharuskan kedua prosedur untuk menggunakan kategori peringkat yang sama.

## 6 Kesimpulan

Pada makalah ini, telah dibahas metode pengukuran kualitas gambar secara subjektif. Metode objektif adalah metode yang praktis apabila dibandingkan dengan metode pengukuran gambar secara objektif. Metode subjektif, meskipun dapat digunakan sebagai patokan bagi pengukuran gambar secara umum,

memiliki keterbatasan yang membuat penggunaannya tidak bisa diterapkan dalam berbagai skenario pengukuran yang bersifat praktis dan *real-time*. Untuk metode pengukuran objektif, beberapa metrik standar telah dibahas. Metrik standar ini memberikan hasil pengukuran yang cepat, namun dalam beberapa hal memiliki kelemahan dari sisi korelasinya dengan persepsi kualitas manusia, berdasarkan kriteria-kriteria evaluasi kinerja model objektif.

## Pustaka

- [1] Yan Ke, Xiaou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 419–426, 2006. doi: 10.1109/CVPR.2006.303.
- [2] Peter G Engeldrum. A theory of image quality: The image quality circle. *Journal of imaging science and technology*, 48(5):447–457, 2004.
- [3] S. B. Patil and S. R. Patil. Survey on approaches used for image quality assessment. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pages 987–991, Aug 2017. doi: 10.1109/ICECDS.2017.8389585.
- [4] Sejal Patil and Shubha Sheelvant. Survey on image quality assessment techniques. *Int. J. Sci. Res*, 4(7):1756–1759, 2015.
- [5] K. Ma, K. Zeng, and Z. Wang. Perceptual Quality Assessment for Multi-Exposure Image Fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, Nov 2015. ISSN 1057-7149. doi: 10.1109/TIP.2015.2442920.
- [6] Maria Persson. Subjective image quality evaluation using the softcopy quality ruler method, 2014. ISSN 1404-6342. Master's theses.
- [7] Andre M. van Dijk, Jean-Bernard Martens, and Andrew B. Watson. Quality assessment of coded images using numerical category scaling. In Naohisa Ohta, Heinz U. Lemke, and Jean Claude Lehureau, editors, *Advanced Image and Video Communications and Storage Technologies*, volume 2451, pages 90 – 101. International Society for Optics and Photonics, SPIE, 1995. doi: 10.1117/12.201231. URL <https://doi.org/10.1117/12.201231>.
- [8] T. Alpert and J. Evain. Subjective quality evaluation - the SSCQE and DSCQE methodologies. *EBU Technical Review*, 271:12–20, May 1997. URL [https://tech.ebu.ch/publications/trev\\_271-evain](https://tech.ebu.ch/publications/trev_271-evain).
- [9] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, 2006. doi: 10.1109/TIP.2006.881959.



- [10] Judith Redi, Hantao Liu, Hani Alers, Rodolfo Zunino, and Ingrid Heynderickx. Comparing subjective image quality measurement methods for the creation of public databases. In Susan P. Farnand and Frans Gaykema, editors, *Image Quality and System Performance VII*, volume 7529, pages 19 – 29. International Society for Optics and Photonics, SPIE, 2010. doi: 10.1117/12.839195. URL <https://doi.org/10.1117/12.839195>.
- [11] Rafał K. Mantiuk, Anna Tomaszewska, and Radosław Mantiuk. Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum*, 31(8):2478–2491, 2012. doi: <https://doi.org/10.1111/j.1467-8659.2012.03188.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2012.03188.x>.
- [12] Mikko Nuutinen, Toni Virtanen, Tuomas Leisti, Terhi Mustonen, Jenni Radun, and Jukka Häkkinen. A new method for evaluating the subjective image quality of photographs: dynamic reference. *Multimedia Tools and Applications*, 75(4):2367–2391, 2016. doi: 10.1007/s11042-014-2410-7. URL <https://doi.org/10.1007/s11042-014-2410-7>.
- [13] Wenhan Zhu, Guangtao Zhai, Menghan Hu, Jing Liu, and Xiaokang Yang. Arrow’s impossibility theorem inspired subjective image quality assessment approach. *Signal Processing*, 145:193 – 201, 2018. ISSN 0165-1684. doi: <https://doi.org/10.1016/j.sigpro.2017.12.001>. URL <http://www.sciencedirect.com/science/article/pii/S0165168417304164>.
- [14] S Opozda and A Sochan. The Survey of Subjective and Objective Methods for Quality Assessment of 2D and 3D Images. *Theoretical and Applied Informatics*, 26(1-2):39–67, 2014.
- [15] ITU-T. International Telecommunication Union. <https://www.itu.int/en/>, 2021.
- [16] VQEG. *Video Quality Experts Group*, 2002. URL <http://www.vqeg.org>.
- [17] Z. Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, March 2002. URL <http://www.cns.nyu.edu/~zwang/>.
- [18] David Sneddon Hands. *Mental processes in the evaluation of digitally-coded television pictures*. PhD thesis, University of Essex, 1998.
- [19] Debarati Kundu, Deepti Ghadiyaram, Alan C Bovik, and Brian L Evans. Large-scale crowdsourced study for tone-mapped hdr pictures. *IEEE Transactions on Image Processing*, 26(10):4725–4740, Oct 2017. ISSN 1941-0042. doi: 10.1109/TIP.2017.2713945.
- [20] Stefan Winkler. *Digital video quality: vision models and metrics*. John Wiley & Sons, 2005.



- [21] Damon M Chandler. Seven challenges in image quality assessment: past, present, and future research. *ISRN Signal Processing*, 2013, 2013.
- [22] VQEG. *Final report from the Video Quality Expert Group on the validation of objective models of video quality assessment – Phase I*. VQEG, March 2000. URL <http://www.vqeg.org>.
- [23] VQEG. *RRNR-TV Group Test Plan Draft Version 1.7h*, June 2004. URL <http://www.vqeg.org>.
- [24] S. Olsson, M. Stroppiana, and J. Baïna. Objective methods for assessment of video quality: State of the art. *IEEE Trans. on Broadcasting*, 43(4):487–495, December 1997.
- [25] Irwan Prasetya Gunawan. *Reduced-Reference Impairment Metrics for Digitally Compressed Video*. PhD thesis, University of Essex, 2006.
- [26] Manish Narwaria, Matthieu Perreira Da Silva, and Patrick Le Callet. HDR-VQM: An Objective Quality Measure for High Dynamic Range Video. *Signal Processing: Image Communication*, 35:46–60, July 2015. ISSN 0923-5965. doi: 10.1016/j.image.2015.04.009. URL <https://www.sciencedirect.com/science/article/pii/S0923596515000703?via%3Dihub>.
- [27] H. Yeganeh and Z. Wang. Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing*, 22(2):657–667, Feb 2013. ISSN 1057-7149. doi: 10.1109/TIP.2012.2221725.
- [28] Tunç Ozan Aydin, Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. Dynamic range independent image quality assessment. *ACM Trans. Graph.*, 27(3):69:1–69:10, August 2008. ISSN 0730-0301. doi: 10.1145/1360612.1360668. URL <http://doi.acm.org/10.1145/1360612.1360668>.
- [29] S. Jia, Y. Zhang, D. Agrafiotis, and D. Bull. Blind high dynamic range image quality assessment using deep learning. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 765–769. IEEE, Sept 2017. doi: 10.1109/ICIP.2017.8296384.
- [30] Stefan Winkler. Visual fidelity and perceived quality: Towards comprehensive metrics. *Proc. SPIE*, 4299, 2001. URL <http://stefan.winkler.net/>. Also downloadable from <http://ltswww.epfl.ch/~winkler>.
- [31] Z. Wang, H. R. Sheikh, and A. C. Bovik. Objective Video Quality Assessment. In In B. Furht and O. Marqure, editors, *Handbook of Video Databases: Design and Applications*, chapter 41, pages 1041–1078. CRC Press, first edition edition, 2003. ISBN 9780849370069.

- [32] Scott Daly. The visible differences predictor: an algorithm for the assessment of image fidelity. *Proc. SPIE - Human Vision, Visual Processing, and Digital Display III*, 1666:2–15, 1992.
- [33] J. Lubin. A human vision system model for objective picture quality measurements. In *Proceedings International Broadcasting Convention*, pages 498–503, Amsterdam, September 12–16 1997.
- [34] Zhou Wang, Ligang Lu, and Alan C. Bovik. Video quality assessment based on structural distortion measurement. *Signal Processing: Image Communication*, 19(2):121–132, February 2004.
- [35] Pina Marziliano, Frederic Dufaux, Stefan Winkler, and Touradj Ebrahimi. A no-reference perceptual blur metric. In *Proceedings of the ICIP*, volume 3, pages 57–60, Rochester, USA, 2002. URL <http://stefan.winkler.net/>.
- [36] Irwan Prasetya Gunawan and Antony Halim. No-reference jpeg image quality assessment using haar wavelet decomposition. *ECTI Transaction On Computer and Information Technology*, 5(2):61–72, November 2011. ISSN 1905-050X. URL <https://www.tci-thaijo.org/index.php/ecticit/issue/view/5253>.
- [37] I. P. Gunawan and M. Ghanbari. Reduced-reference video quality assessment using discriminative local harmonic strength with motion consideration. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(1): 71–83, Jan 2008. ISSN 1051-8215. doi: 10.1109/TCSVT.2007.913755.
- [38] Irwan Prasetya Gunawan and Mohammed Ghanbari. Efficient reduced-reference video quality meter. *IEEE Transactions on Broadcasting*, 54(3):669–679, September 2008. ISSN 0018-9316 (Print), 1557-9611 (Online). doi: 10.1109/TBC.2008.2000734. URL <https://ieeexplore.ieee.org/document/4550733/>.
- [39] A. J. Viera and J. M. Garrett. Understanding interobserver agreement: The kappa statistic. *Society of Teachers of Famili Medicine*, 2005.