Recommendation

ITU-T F.740.4 (09/2023)

SERIES F: Non-telephone telecommunication services

Multimedia services

Metadata for image aesthetics assessment with aesthetic attributes in mobile terminal computational photography systems

ITU-T F-SERIES RECOMMENDATIONS

Non-telephone telecommunication services

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T F.740.4

Metadata for image aesthetics assessment with aesthetic attributes in mobile terminal computational photography systems

Summary

Image aesthetics assessment (IAA) aims to evaluate whether an image conforms to the aesthetic preferences of a potential audience. Recommendation ITU-T F.740.4 defines metadata for image aesthetics assessment with aesthetic attributes in a mobile terminal computational photography system. The metadata is divided into three dimensions according to the key roles: photographer, camera and viewer, and provides non-redundant, fundamental and representative aesthetic attributes of each dimension. The role-based metadata defined in this Recommendation can be used to guide the construction of IAA datasets, and to provide multiple aesthetic attributes evaluation for IAA. Moreover, sample collection requirements and qualifications for annotators are recommended to guide high quality data construction.

History *

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T F.740.4	2023-09-13	16	11.1002/1000/15614

Keywords

Image aesthetics assessment, metadata, subjective evaluation.

---

* To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T F.740.4

Metadata for image aesthetics assessment with aesthetic attributes in mobile terminal computational photography systems

1 Scope

This Recommendation defines metadata for image aesthetics assessment (IAA). The scope of this Recommendation includes:

1) The structure of metadata for IAA.

2) The dimensions of metadata for IAA.

3) Requirements for sample collection.

4) Requirements for labelling equipment and environment.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

None.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following term defined elsewhere:

3.1.1 metadata [b-ITU-T X.1255]: Structured information that pertains to the identity of users, systems, services, processes, resources, information or other entities.

3.2 Terms defined within this Recommendation

This Recommendation defines the following terms:

3.2.1 aesthetic attributes: Attributes that affect the human perception of image aesthetics, typically including both subjective and objective attributes.

3.2.2 image aesthetics assessment: The process of evaluating the aesthetic quality of an image using computational techniques.

3.2.3 sample collection: The process of acquiring the images involved in the construction of the dataset.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

IAA   Image Aesthetics Assessment

5   Conventions

In this Recommendation:

–   The keywords "is required to" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

–   The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

–   The keywords "should" indicate a requirement that is not mandatory. It implies that a certain course of action is preferred and expected, but that there may be situations where it is not practical or feasible to follow the recommendation.

6   Overview

This Recommendation standardizes metadata related to image aesthetics assessment (IAA), the metadata is divided into three dimensions according to the key roles: photographer, camera and viewer, and provides non-redundant, fundamental and representative aesthetic attributes of each dimension. Three aspects are then recommended to ensure the quality of the dataset constructed by metadata guidance: firstly, sample collection requirements are recommended to ensure comprehensive samples. Secondly, once the samples that meet the requirements are collected, they will be annotated by annotators, thus the qualifications for annotators are recommended to ensure that the annotators have the relevant skills to provide high-quality annotations. Thirdly, the labelling equipment and environment also affect the annotation quality, and some requirements are recommended.

7   The dimensions of metadata

According to the logical structure, shown in Figure 1, IAA metadata for the mobile terminal computational photography system is divided into the following:

–   **Photographer** - Metadata mainly includes some aesthetic attributes manipulated by the photographer, which are related to photography techniques that photographers often use, such as image composition, colour, camera angle, etc. The image aesthetics can be evaluated indirectly from these attributes.

–   **Camera** - Metadata mainly includes some aesthetic attributes that affect image quality, such as image noise, exposure, sharpness, etc. The image aesthetics can be evaluated indirectly from these attributes.

–   **Viewer** - The aesthetic evaluation of an image ultimately depends on its viewer, and some attributes that are related to the viewer's cultural customs or content preferences. For example, if a person likes horses, he or she probably also likes an image of horses. The image aesthetics can be evaluated directly from these attributes.
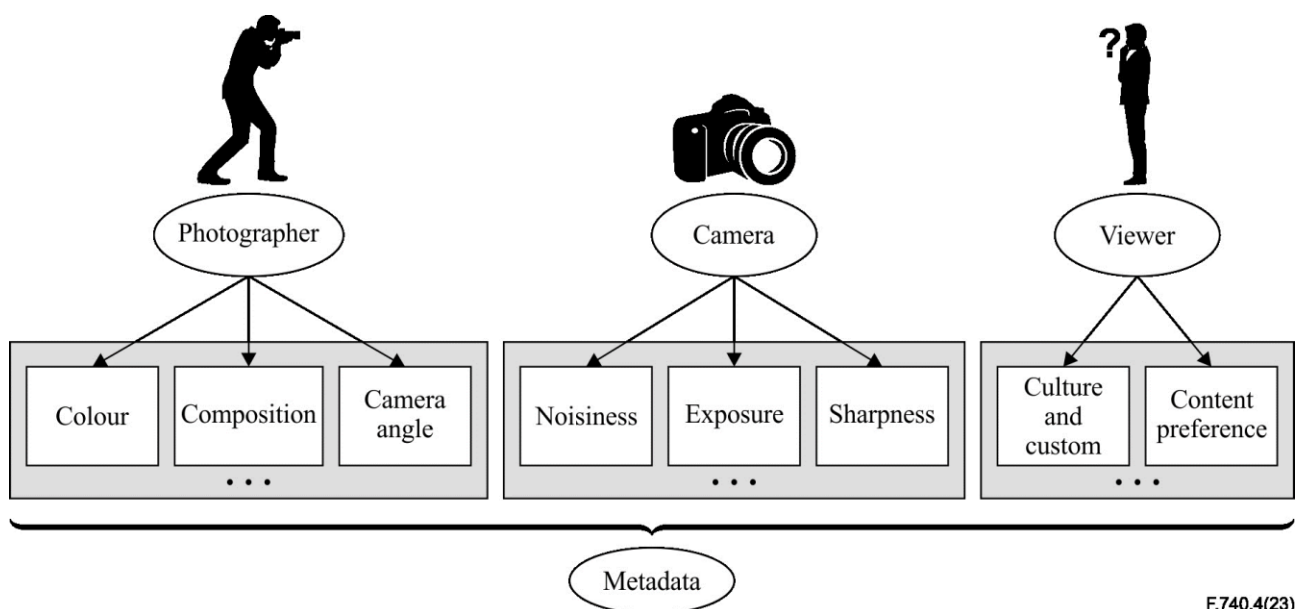
Figure 1 – The logical structure of metadata

8 Metadata for image aesthetics assessment

The metadata of each dimension is composed of various attributes, and the annotation of each attribute is divided into multiple levels to distinguish the aesthetics.

8.1 Metadata of photographer

The photographer can adjust some key aesthetic attributes to enhance the aesthetics of the subject. In order to facilitate the evaluation of image aesthetics from a more professional perspective, the aesthetic attributes listed in Table 1 are mainly related to photography techniques. Each annotation level is divided into several levels, for example, 1~3 means the attribute has 3 different annotation levels. The level is recommended to be divided into: level 1: Inappropriate, level 2: Normal and level 3: Appropriate, to rate the appropriateness of these aesthetic attributes.

Table 1 – Metadata of photographer

Attributes	Description	Level	Type
Composition	The spatial relationship between the subjects, the position of each visual element.	1~3	integer
Colour	The type and intensity of the colour.	1~3	integer
Camera angle	Shooting height, direction and distance.	1~3	integer
Light and shadow	The combination of light and shadow when shooting.	1~3	integer
Special skills	The special techniques used to express the aesthetics of the subject when shooting.	1~3	integer

8.2 Metadata of camera

The camera mainly affects the image quality through some attributes, and further affects the aesthetics of the image. As shown in Table 2, each annotation level is divided into several levels, for example, 1~7 means the attribute has seven annotation levels. Sharpness, exposure, white balance, and saturation are rated with 5 levels, which are used to evaluate whether these attributes are appropriate during shooting. The level of sharpness is divided into: level 1: Severely blurred,

level 2: Slightly blurred, level 3: Normal, level 4: Sharp and level 5: Extremely sharp. The level of exposure is divided into: level 1: Severely underexposed, level 2: Slightly underexposed, level 3: Normal, level 4: Slightly overexposed and level 5: Severely overexposed. The level of white balance is divided into: level 1: Too warm, level 2: Warm, level 3: Normal, level 4: Cold, and level 5: Too cold. The level of saturation is divided into: level 1: Too dim, level 2: Dim, level 3: Normal, level 4: Vivid, and level 5: Too vivid. Noisiness and artifact are rated with 3 levels: level 1: Severely, level 2: Slightly and level 3: None. Focus is rated with 3 levels: level 1: Severely defocused, level 2: Slightly defocused, level 3: Normal.

Table 2 – Metadata of camera

Attributes	Description	Level	Type
Sharpness	The clarity of image boundaries and details.	1~5	integer
Exposure	The amount of light that makes an image clear.	1~5	integer
White balance	The same general appearance as the colours in an original scene.	1~5	integer
Saturation	The vividness of the colour.	1~5	integer
Noisiness	The degree of random variation of brightness or colour information; The isolated pixels that produce strong visual effects.	1~3	integer
Artifact	The degree of anomalies apparent or non-existent elements during visual representation.	1~3	integer
Focus	The accuracy of the object and range of focus during shooting.	1~3	integer

8.3 Metadata of viewer

In the process of IAA, the image is finally evaluated by the viewer, so the viewer can directly determine whether the image is attractive. Unlike the previous dimensions of "camera" and "photographer", this dimension collects some inherent attributes that do not change with different application scenarios. The aesthetic attributes of viewer should be determined according to specific communities and the application scenario, which mainly includes two dimensions: content preference, and culture and custom, as shown in Table 3. For example, some people like "green", while others dislike "green", thus it is necessary to add annotation information, e.g., "Do you like the green colour?" to describe the preference. Each annotation level is divided into 2 Boolean levels, e.g., Level 0: Like, Level 1: Dislike.

Table 3 – Metadata of viewer

Attributes	Description	Level	Type
Content preference	Preferences for specific subjects.	0, 1	bool
Culture and custom	Preferences for specific culture and custom.	0, 1	bool

9 Requirements for sample collection

When constructing a dataset, the sample images should be collected in a comprehensible scene and should also meet the sample coverages of each attribute. This Recommendation specifies the scene and sample coverages for each attribute in the sample collection, as well as some special coverages, including content size, shot distance, camera facing, etc., as shown in Table 4.

Table 4 – Requirements for sample collection

Types	Scene	Coverages
Composition	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd)	Composition type (rule of thirds, centre, symmetry, diagonals and triangles, horizontal line), content size (small, medium, large), shot distance (long, medium, close)
Colour	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd), night scene (indoor, outdoor)	Colour combination (complementary colours, monochromatic colours, analogous colours)
Camera angle	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd)	Angle and position type (high, eye-level, low), content size (small, medium, large), shot distance (long, medium, close), camera facing (front, rear)
Light and shadow	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd), night scene (Indoor, outdoor)	Light type (front, side, back, reflection), camera facing (front and rear), the state of motion (static, slow, fast)
Special skills	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd), night scene (indoor, outdoor)	Skill type (virtual reality, panorama, rotation, achromatism), content size (small, medium, large), shot distance (long, medium, close), camera facing (front and rear)
Sharpness	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd), night scene (indoor, outdoor)	Special sharpness (background with defocus, focus), content size (small, medium, large), shot distance (long, medium, close), camera facing (front and rear), the state of motion (static, slow, fast)
Exposure	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd), night scene (indoor, outdoor)	Special exposure (day overexposed, day underexposed, night overexposed, night underexposed, night normal exposed), camera facing (front and rear), tone (high-key, mid-tone, low-key)
White balance	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd), night scene (indoor, outdoor)	Light source (uniform, non-uniform, single, multiple), camera facing (front and rear), colour or texture type (foreground: mixed colour, orange, yellow, white, blue; background: monochromatic colour, pattern, wood grain)
Saturation	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd), night scene (indoor, outdoor)	Colour type (pure colours: bright red, bright green, gray; mixed colours: white gray, magenta, pink, yellowish brown), camera facing (front and rear)
Noisiness	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd), night scene (indoor, outdoor)	The state of motion (static, slow, fast), camera facing (front and rear)

Table 4 – Requirements for sample collection

Types	Scene	Coverages
Artifact	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd), night scene (indoor, outdoor)	Artifact type (image ghost, image moire, image vignetting), the state of motion (static, slow, fast), camera facing (front and rear), content type (hair, striped clothing, checked clothing, brick walls, fences, electronic device screens)
Focus	Natural scenery, urban scenery (indoor, outdoor), person (single, multiple, crowd), night scene (indoor, outdoor)	Special sharpness (background with defocus, focus), camera facing (front and rear), the state of motion (static, slow, fast), shot distance (long, medium, close), content size (small, medium, large)

10 Requirements for labelling equipment and environment

It is important to specify the specifications of the equipment and the environment when manually annotating image aesthetic datasets, as shown in Table 5. By doing so, it is possible to ensure that the annotations are accurate, consistent, and reliable, and that the results based on the annotations are meaningful and statistically robust.

Table 5 – Requirements for labelling equipment and environment

Types	Requirements
Labelling software	Multiple images should be able to be conveniently and continuously rated with scores.
Storage format	It is recommended to use a simple format, such as XML or JSON.
Management system	A data labelling platform should include an integrated management system to manage projects, data, and users, and should enable project managers to track project progress and user productivity and communicate with annotators about mislabelled data, as well as sampling for quality assurance.
Device image format	The format of the images (e.g., JPEG, PNG, TIFF) should be specified.
Device colour space	The colour space should be specified to ensure consistency and accuracy in the labelling of the images. Common colour spaces include RGB, grayscale, and CMYK.
Device resolution	This Recommendation states that the device used for labelling images should have a display with a high resolution to ensure that the image is displayed in its entirety.
Device brightness	The brightness level of the device should ensure that it is within the appropriate brightness range.
Device hue	The device should have accurate colour representation and good colour balance. It is recommended to use a device with a colour calibration tool to ensure accurate colour representation.
Device refresh rate	This Recommendation states that the device is able to display images smoothly while being used for labelling. A higher refresh rate can further improve the smoothness of the display.
Device contrast	The device should have a suitable contrast ratio to ensure good

Table 5 – Requirements for labelling equipment and environment

Types	Requirements
	image quality and visibility of details in the images.
Workstation settings	The workstation should be designed ergonomically, with the monitor placed at a comfortable viewing height and angle, and the keyboard and mouse positioned to encourage good posture. The monitor should also be positioned to reduce glare and reflections.
Ambient temperature	The temperature in the work area should be maintained at a comfortable level to ensure a conducive working environment.
Lighting	The lighting in the work area should be adequate for precise image labelling, but not so bright that it causes eye strain. A well-lit workspace with adjustable brightness and colour temperature can ensure optimal lighting conditions.
Rest breaks	Annotators should take regular rest breaks to avoid eye strain, fatigue, and repetitive stress injuries, during which time the labeller should look away from the monitor, stretch, and rest the eyes.
Training and supervision	Annotators should receive training on the labelling process and the use of the equipment, as well as ongoing supervision to ensure that the labelling process is accurate and consistent.

Appendix I

Supplements for the qualification evaluation of annotators

(This appendix does not form an integral part of this Recommendation.)

I.1 The qualification evaluation of annotators

Since the annotation of image aesthetics is highly subjective, the evaluation of aesthetics is influenced by some background information of the annotators. Therefore, the qualifications of the annotators should be evaluated to ensure that the quality of their annotations is high enough. However, this is not mandatory and can be specifically formulated according to the actual application scenario. At the same time, the evaluation items need to be agreed upon by the annotators. The qualifications are shown in Table I.1.

Table I.1 – Qualification evaluation for annotators

Types	Qualifications
Attention to detail	Annotators should be meticulous and detail-oriented, as accuracy is crucial when annotating aesthetic image datasets. They should have the ability to identify and label important features in the images, such as colour, texture, and patterns, with precision and consistency.
Aesthetic sensitivity	Annotators should possess a sense of aesthetics as they will evaluate images based on their visual appeal. This requires a level of subjective judgment and the ability to identify visually attractive features.
Visual memory	Annotators should have a good visual memory as they need to remember and compare the details of different images while labelling them.
Technical proficiency	Annotators should be proficient with the software and tools used for the annotation process, including image annotation software and the hardware used for display and input. This will enable them to work efficiently and accurately, and minimize the risk of errors in the labelling process. Additionally, proficiency with the tools may also allow them to identify potential technical issues or glitches that could affect the accuracy or efficiency of the annotation process.
Consistency	Annotators should strive to maintain consistency in their labelling, applying the same standards and criteria to each image they annotate. This is particularly important when working in a team, as consistent labelling by all team members is essential for the accuracy and reliability of the annotations.
Communication skills	Annotators should have enough communication skills as they may need to ask questions or provide feedback to other team members, such as project managers or supervisors. Clear communication helps to ensure that everyone is on the same page and that the labelling process proceeds smoothly and accurately.
Patience and focus	Annotators should have the patience and focus to work on repetitive tasks for extended periods of time since annotating a large number of images can be a time-consuming process.
Receptiveness to training	Annotators should be willing to receive training and be open to feedback, as they may need to learn new skills or improve their performance during the course of the project.

Bibliography

[b-ITU-T X.1255]   Recommendation ITU-T X.1255 (2013), *Framework for discovery of identity management information.*

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	**Non-telephone telecommunication services**
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems