Research on Image Content Metadata Framework and Standard Development (Postprint)

Authors: Zhang Chi, Huang Jing, Zhang Pengzhou, Wu Guowei

Date: 2023-10-08T00:00:00+00:00

Abstract

Images have become one of the most efficient communication media connecting media, brands, and consumers. Image feature description constitutes an important foundation for image retrieval, recommendation, and exchange. Based on an analysis of the current state of image applications and image metadata standards, this article proposes an image content metadata framework that emphasizes the description of image content and semantic features, and based on this framework, completes the formulation of the national standard 'Chinese News Image Content Description Metadata Specification', thereby filling the gap in domestic image metadata standards, helping to improve the efficiency and quality of image annotation, and better meeting the needs of image acquisition and transactional exchange.

Full Text

Research on Image Content Metadata Systems and Standard Development

Abstract: Images have become one of the most efficient communication media connecting media, brands, and consumers. Image feature description is a critical foundation for image retrieval, recommendation, and exchange. Based on an analysis of current image applications and image metadata standards, this paper proposes a framework for an image content metadata system that focuses on describing image content and semantic features. Building upon this framework, the national standard "Specification for Content Description Metadata of Chinese News Images" was developed, filling a gap in domestic image metadata standards, helping to improve the efficiency and quality of image annotation, and better meeting the needs of image acquisition and transactional exchange.

Keywords: metadata; image content metadata; national standard

Images play important roles in news, publishing, advertising, marketing, and other fields. Particularly in the context of convergent media development, images—with their advantages of high information density, strong visual appeal, and small data size—are playing an increasingly significant role in applications such as WeChat-Weibo-client platforms and premium content production. Discovering and acquiring high-quality images is the core aspect of image applications, and image feature representation is the cornerstone supporting this core process. In recent years, deep learning technology has made breakthrough progress in image processing fields such as handwritten character recognition, face recognition, image-based search, and image segmentation [1]. On the other hand, text-based image metadata remains the primary means of image feature description, and with the continuous development of text processing and natural language processing technologies, its research and application have broad prospects for development.

1. Research Foundation

Research on image applications, composition, and editing deepens our understanding of images from different perspectives and provides an important foundation for constructing an image content metadata system.

1.1 Image Applications

Current mainstream image libraries can be roughly categorized into several types. First are large comprehensive commercial image libraries, such as Visual China and East Photo, which have numerous professional contracted photographers and primarily provide high-quality editorial images, creative images, and micro-video resources to institutional users. Second are media organization image libraries, such as the China Global Photo Alliance and various newspaper image libraries, which focus on editorial images and each possess distinctive resources. Third are design material libraries, such as Quanjing, Yitu, Nipic, Lanrentuku, and Paixin, which provide various design assets. Fourth are vertical domain-focused libraries specializing in e-commerce, automotive, geography, photography, web materials, etc. Additionally, there are social image libraries centered on personal sharing and communication. Regardless of type, the most critical function is to serve as a bridge between image creators and consumers, with image discovery being the value manifestation of this bridge.

Images must be discovered and circulated to realize their value, which requires proper description and storage. Regarding the current status of image metadata applications, we conducted field research at several large-scale commercial and media organization image libraries in China. The findings revealed that: (1) as image volumes grow and circulation speeds increase, image retrieval becomes increasingly difficult; (2) existing image metadata standards cannot adequately meet image indexing and retrieval needs, particularly lacking metadata related to content semantics; (3) current image annotation practices are chaotic, with

non-uniform metadata standards, widespread use of proprietary custom metadata, and a lack of quality control mechanisms in the annotation process; and (4) there is urgent demand for unified image metadata specifications.

Based on these findings, this paper proposes an image content metadata system oriented toward thematic and semantic description of images to enhance semantic completeness in annotation, improve standardization and compatibility of annotation results. Through comparative analysis of representative image library systems, image retrieval functions can be primarily classified as follows:

Keyword search: Most commonly used, generally highly relevant to image themes, objects, people, locations, and events.

Category-based search: Category divisions vary significantly. In addition to conventional knowledge domain classification methods, shallow tag-based thematic classification is increasingly adopted.

Thematic event search: Widely used in news image organization and queries.

Image attribute filtering: Improves retrieval accuracy and helps quickly locate images. Common attributes include image source (individual/organization), licensing status, color, size, format, and people information (number, age, gender, ethnicity), as well as background settings.

Image recommendation: 主要包括热门推荐、编辑推荐、相似推荐、个性化推荐等。

Refined search: Main implementation methods include: searching within results; statistical inverted indexing of tags from previous search result sets for user selection; and related tag recommendations.

Analysis of these retrieval functions reveals several key insights: (1) using text to find images remains the primary retrieval method—keywords, categories, thematic events, filtering attributes, and even recommendations all depend on textual annotation information; (2) the number of descriptive dimensions for image features is increasing, with enhanced descriptions adding dimensions and improving precision (e.g., evolving from simple captions to detailed keyword lists) to address the need for rapid location of target images from massive collections; (3) existing image metadata standards inadequately meet practical application needs, with widespread use of proprietary custom metadata items reflecting both common and personalized requirements; and (4) the value of image content and semantic features is becoming prominent—visual content, expression form, and thematic concepts constitute the soul of images, with nearly all retrieval functions related to image content.

1.2 Image Composition

Composition refers to using visual features to reproduce real-world objects in two-dimensional space, conveying additional information through frame construction to reflect the creator's understanding and emotions, while serving to

highlight subjects, attract attention, simplify complexity, and create harmonious balance. Composition emphasizes arranging people, scenery, and objects within the frame to achieve optimal layout, employing visual elements such as points, lines, shapes, lighting, and color for aesthetic appeal. The purpose of composition is to convey information, express themes, and communicate the creator's cognition and feelings.

From a compositional perspective, an image primarily comprises three elements: subject, supporting objects, and environment. The subject is the main subject of representation and often serves as the structural and visual center of the image. Supporting objects help express the subject's characteristics and connotation by forming certain narratives with the subject. The environment, divided into foreground and background, enhances the subject and narrative. Properly managing the relationships among subject, supporting objects, and environment is key to theme expression and image quality evaluation. Additionally, lighting, tonal quality, and photographic techniques are indispensable in image creation.

1.3.1 Photojournalism

The *Photo Editing Handbook* [3] clearly defines photojournalism as a reporting form combining images and text in media, encompassing various forms of photographic reporting in media. Creative photography and feature news reporting are different manifestations of photojournalism, which differs from propaganda photography (public relations photography) and pictorial photography. Imagery and newsworthiness are the two fundamental characteristics of photojournalism. Excellent photojournalism should possess historical value, social value, psychological impact, and aesthetic value. A photograph provides readers with not only visual content itself but also implicit information such as themes, visual aesthetics or impact, emotions, and artistic conception expressed through visual content.

1.3.2 Photo Evaluation

The early evaluation criteria for news photos were "novelty, truthfulness, vitality, emotion, and meaning." With the continuous development of media, evaluation standards for photojournalism have evolved to include technical criteria, information transmission, aesthetic criteria, and communication effects. Technical criteria (including exposure, color temperature, depth of field, focus, etc.) form the foundation, with emphasis on whether the photo accurately and richly conveys semantic information and, more importantly, whether it achieves good communication effects (both content and form significantly impact communication effectiveness).

1.3.3 Photo Caption Writing

Photo captions should explain both easily noticeable elements and important yet less obvious details while avoiding bias and subjective evaluation. Single-

image captions are typically completed in two sentences: the first describes what is happening in the image, specifying time, location, people, and a brief event description; the second provides relevant background. Group photo caption writing methods include directly stating event information, selecting from multiple aspects of an event, or emphasizing background introduction. Generally, single-image captions focus on identifying people, scenes, and events, while group captions emphasize background introduction and in-depth description.

Evidently, whether regarding photojournalism characteristics, photo evaluation criteria, or caption writing, the focus extends beyond the visual content itself. Therefore, multi-dimensional feature description encompassing technical, aesthetic, event, background, and important detail perspectives is significant.

2. Image Content Metadata System

2.1 Hierarchy of Image Metadata

Metadata is data about data. In the digital library domain, metadata is divided into three types: descriptive, administrative, and structural. The *IPTC Photo Metadata* standard [4], widely adopted internationally, divides image metadata into descriptive and technical categories. This paper classifies image metadata into four layers based on semantic abstraction level: physical, logical, content, and thematic layers, as shown in Figure 1 [Figure 1: see original paper].

Physical layer metadata primarily includes digital image file attributes, technical parameters during image capture, and low-level visual features. Logical layer metadata includes image application attributes and licensing information. The content and thematic layers, built upon the physical and logical layers, describe the visual content and thematic information presented in images. The content layer focuses on visible objects in the frame, while the thematic layer emphasizes concepts and ideas expressed through visual content. These four layers can be further grouped into attribute metadata (physical and logical layers) and content metadata (content and thematic layers). This paper proposes an image content metadata system targeting the latter.

2.2 Image Content Metadata System

The *IPTC Photo Metadata* standard focuses on descriptive metadata, primarily covering logical and content layers. It has released multiple versions with unstructured relationships between metadata elements and does not address thematic layer metadata. The EXIF standard mainly emphasizes physical layer metadata. The *Chinese News Information Markup Language* [5] primarily targets news manuscripts without detailed specifications for image feature description. Combined with previous analysis, constructing a content- and semantics-oriented image metadata system and developing standard specifications are needed and valuable for image applications, image characteristics, and practical situations.

How should image content be defined? When viewing an image, readers intuitively perceive objects, colors, and composition, which fall within the scope of image content. Simultaneously, the mood, ideas, and emotions that these visual elements evoke through perception, association, and appreciation also constitute important components of image content.

How should image content be described? Building upon the research foundation described earlier and adopting a combined bottom-up and top-down approach, this paper ultimately forms a tripartite architecture. Image content metadata includes three aspects: visual form, technical execution, and thematic concept. Visual form is the visual representation, technical execution is the creative method, and thematic concept is the image's theme and soul. Visual form and technical execution serve the thematic concept, which is expressed through the former two. These three aspects are not isolated but complementary and integrated into an organic whole.

Specifically, "visual form" refers to the main objects in an image and their characteristics, as well as environmental information. "Technical execution" refers to information related to photographic techniques. "Thematic concept" refers to image themes, semantics, and knowledge domain classification. These three dimensions contain several sub-dimensions, as shown in Figure 2 [Figure 2: see original paper]. Since most dimensions are self-explanatory, detailed descriptions of each are not provided here.

3. Standard Development

The *Specification for Content Description Metadata of Chinese News Images* was approved by the National Standardization Administration in September 2014. As a project team member, the author was primarily responsible for drafting the standard document. The development received strong support and guidance from experts in standardization, media organizations, commercial image libraries, academic institutions, and relevant technology companies. Based on the image content metadata system and through repeated research, drafting, review, feedback, and revision, the draft for approval has been completed and submitted to the National Standardization Administration.

3.1 Development Principles

Applicability: With business requirements as the starting point, the standard emphasizes content and semantic feature description, reduces hierarchy, and increases the proportion of controlled vocabularies used during annotation.

Coordination: For metadata elements with identical or similar meanings in existing relevant standards, consistent naming was adopted to avoid confusion during multi-standard application.

Professionalism: Targeting image content and semantic feature description, metadata elements focus on three perspectives: visual content, thematic con-

cept, and photographic technique, with specific guidance from multiple domain experts.

The *Specification for Content Description Metadata of Chinese News Images* defines a metadata element set for describing content and semantic features of Chinese news images from two application perspectives: news editing images and creative illustration images, applicable to image data collection, editing, storage, publication, retrieval, exchange, and other processing stages.

The metadata element set is divided into three parts: common metadata, news editorial image metadata, and creative image metadata, comprising 40 metadata elements total. Among these, "people information" and "photographic technique" contain secondary elements, as shown in Figure 3 [Figure 3: see original paper].

Comparing Figures 2 and 3 reveals that the standard adds several logical layer metadata elements, such as title, caption, capture time, capture location, and identifier. These were added for application convenience as they are required in practice. Common metadata elements apply to all image categories, so format, color, and shot size were moved from "photographic technique" to common metadata. The standard does not define metadata for image copyright, as copyright is not closely related to image content; it is recommended to directly reference existing copyright-related standards in image applications.

The standard specifies only 8 mandatory elements, concentrated in the common and editorial image metadata sections. It provides 22 controlled vocabularies for annotating 20 metadata elements. Detailed definitions and specifications are available in the standard document.

3.3.1 Emphasis on Standardization in Annotation

Using controlled vocabularies helps improve annotation accuracy and consistency, and the standard supports controlled vocabulary extension. For metadata elements using free text annotation, custom annotation specifications are also recommended.

3.3.2 Emphasis on Completeness in Annotation

The standard specifies few mandatory metadata elements to better meet the need for rapid image publication. A multi-level annotation approach is recommended, combining coarse-grained and fine-grained annotation. Additionally, applications can refine annotation requirements for certain metadata elements—for example, adding secondary metadata for "brand" in fashion images or "breed" in animal images under the "main subject" element.

3.3.3 Combining Automatic and Manual Annotation

Utilizing computer image processing technology enables automatic annotation of elements such as "color," "people count," "color tone," and "tonal quality."

Automatically extracting keywords from image descriptive text can improve annotation efficiency for dimensions like “theme,” “location,” and “time.”

The fundamental goal of standard development is to better meet image retrieval and exchange needs. Multi-dimensional annotation result sets also provide an important foundation for image content analysis, resource association and aggregation, and similar image recommendation.

Image content feature description plays a crucial role in image applications. With deep learning representing breakthrough progress in computer vision processing technology, how to combine text-based image feature description with automatic processing technology to better meet image business requirements across multiple scenarios represents important research and application value.

References [1] Liu Jianwei, Liu Yuan, Luo Xionglin. Research Progress on Deep Learning. Computer Application Research. Vol. 31, No. 7. 2014.7. [2] China Internet Network Information Center. 40th Statistical Report on China’s Internet Development. 2017(7). [3] Ren Yue, Zeng Huang. *Photo Editing Handbook* (4th Edition). China Photography Publishing House. 2015(9). [4] IPTC Photo Metadata Standard. http://www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata. [5] National Technical Committee for Chinese News Information Standardization. National Standard GB/T 2009-2013 *Chinese News Information Markup Language*.

(Author affiliation: Xinhua News Agency Communication Technology Bureau)

Note: Figure translations are in progress. See original paper for figures.