**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC 1/SC 29/WG 11 N15087**

**Geneva, CH – February 2015**

|  |  |
| --- | --- |
| **Source** | **Requirements** |
| **Status** | **Approved** |
| **Title** | **Data Requirements for Compact Descriptors for Video Analysis in Search and Retrieval Applications** |

# Introduction

MPEG will make a public Call for Proposals on Compact Descriptors for Video Analysis (CDVA), addressing visual search and retrieval applications in video content. This document presents an outline of the data required for the evaluation framework [1] of this upcoming call and requests interested experts to contribute such data. Both professional video content and user-generated content is applicable.

# Data for Compact Descriptors for Video Analysis for Search and Retrieval Applications

## Items of Interest

The items of interest in the upcoming Call for Proposals are planar or non-planar, rigid or partially rigid, textured or partially textured objects or scenes.

Faces and object classes (e.g. "animal", "vehicle", etc.) are not included and are outside the scope of the Call.

## Item Categories

The items of interest are split into three broad categories.

1. Large objects like building facades, landmarks, etc.

These are generally large objects which will appear as backgrounds obstructed by varied foreground objects.

1. Smaller objects like paintings, books, statues, signs, etc.

These are generally smaller objects which may appear on varied backgrounds and may also be obstructed by other foreground objects.

1. Scenes like interior scenes, natural scenes, multi-camera shots, etc.

These are generally camera recordings showing the same things or actions from different viewpoints.

## Video Clips

The evaluation framework for the Call requires

* For each of the three item categories (Large objects, Smaller objects, Scenes), at least 250 items of interest (e.g. "Big Ben", "Mona Lisa", etc.), giving a total of at least 750 items across all categories.
* For each item of interest, 4 – 8 video clips showing the item at different scales and/or illuminations and/or viewpoints and/or recorded at different times and/or at different video qualities, etc.

Each video clip can last from 1 second to any duration and should

* depict the item of interest from start to finish;

OR

* depict the item of interest for most but not all frames (e.g. a panning shot where the object comes into view after the start and then moves out of view before the end);

OR

* be accompanied by some frame number / time stamp annotation for the item of interest in the case of long videos where the item of interest appears only briefly at some point. This annotation will allow the CDVA experts to easily find the item of interest in the long video.

## Video Clip Qualities and File Formats

There are no special restrictions, as the evaluation framework for the Call requires a broad range of video resolutions and qualities.

There are no special restrictions on file formats and any common file format may be submitted.

## Content Difficulty

The evaluation framework for the Call requires a broad range of matching difficulties. Extreme scale or viewpoint changes can be used to test the performance limits of proposed technologies, but should not form the main corpus of the evaluation framework data. Content contributors should keep in mind that the item of interest should be easily recognizable by a casual human observer.

## Required Metadata

Video clips should be annotated as shown in the example below to allow CDVA experts to organize them in the evaluation framework. This information can be provided as an Excel table, comma-separated values, or other suitable means.

**Video Clip Name Item Category Item Instance Item Name Contributor Copyright/Licensing**

Xhhsiebw.mp4 Large Objects 1 Big Ben <Name> <…>

Skjlkshu6fgsg.avi Large Objects 2 Big Ben <Name> <…>

………

Jhsj73bd6.mpegts Large Objects 7 Big Ben <Name> <…>

jshua.mpg Smaller Objects 1 Mona Lisa <Name> <…>

………

Additionally, only in the case of long videos where the item of interest appears briefly at some point, the long video should be accompanied by some frame number / time stamp annotation for the item of interest, so that CDVA experts may find it.

## Optional Metadata

The following information is optional and desired if available:

Scene Description/Tags

Object Description/Tags

Location Description/Tags

Location Coordinates

Cardinal Direction

Date

Time

Camera Orientation

Camera model/type

Lens

Aperture

ISO

Shutter Speed

Lighting conditions

Special camera conditions

Video Resolution

Frame Rate

Video File Format

Bitrate

Number of frames/Duration

# Additional Information and Content Submission

Content contributors may obtain additional information and clarifications by sending an email to the CDVA reflector (cdvs-tnt@listserv.uni-hannover.de) or the AHG chairs.

An ftp site has been established for content and annotation upload. Content contributors should send an email to the CDVA reflector to obtain an account. Each class of video should be uploaded to a directory specifcally created with a name such as NameOfContributor\_TypeOfVideo.

# References

1. N15088, Compact Descriptors for Video Analysis for Search & Retrieval: Draft Evaluation Framework, Feb. 2015, Geneva, CH