**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11/N15339**

**July 2015, June, Warsaw**

**Title: Call for Proposals for Compact Descriptors for Video Analysis (CDVA) – Search and Retrieval**

**Source Requirements Subgroup**

**Status: Approved**

# 1. Introduction

MPEG is planning standardizing technologies that will enable efficient and interoperable design of visual analysis applications, called Compact Descriptors for Video Analysis (CDVA, see [1] for an introduction). MPEG identified several functionalities that are of interest to industry and collected relevant requirements. The three applications of particular interest are: (1) Search and Retrieval (2) Object and Event Detection and (3) Scene Classification.

This call addresses descriptor technology for search and retrieval applications, i.e. for visual content matching in video. Visual content matching includes matching of views of large and small objects and scenes, that is robust to partial occlusions as well as changes in vantage point, camera parameters, and lighting conditions. The objects of interest comprise planar or non-planar, rigid or partially rigid, textured or partially textured objects, but exclude the identification of people and faces. The databases can be large, for example broadcast archives or videos available on the internet. More details about the intended use cases can be found in [4]. The industry thus needs video descriptors that enable performing this task with smaller descriptor sizes and shorter runtimes as compared to application enabled by single-frame (still image) descriptors (e.g. CVDS) in the video domain.

There are a number of component technologies that are useful for visual search, including format of visual descriptors, descriptor extraction process, as well as indexing, and matching algorithms. As a minimum, the format of descriptors as well as parts of their extraction process should be defined to ensure interoperability.

It is envisioned that a standard for compact descriptors for video analysis for search and retrieval applications will:

* enable design of interoperable object instance search applications
* minimize the size of video descriptors
* ensure high matching performances of objects (in terms of accuracy and complexity)
* enable efficient implementation of those functionalities on professional or embedded systems

It is envisioned that such a standard will provide a complementary tool to the suite of existing MPEG standards, such as MPEG-7 Compact Descriptors for Visual Search.

The requirements for the technology that MPEG intends to standardize are listed in Appendix A of this document.

# 2. Timetable and Procedures

**2.1. Overview**

A timetable for the Call for Proposals relative to specific MPEG meetings is given in the following table. Details of these meetings are available at http://www.chiariglione.org/mpeg/meetings.

|  |  |  |
| --- | --- | --- |
| **MPEG Meeting #** | **Date** | **Action** |
| 112 | 2015/06/22-26 | CfP issued |
|  | 2015/08/15 | Datasets and Evaluation Framework released (see distribution details) |
| 114 | 2016/02/20-21 | Initial evaluation of proposals and assignment of cross-checks |
| 114 | 2016/02/22-26 | Detailed evaluation |
| 115 | 2016/05/30-2016/06/03 | Crosscheck and additional evaluation as required |
| 116 | 2016/10/17-21 | WD |
| 117 | 2017/01/16-21 | CD |
| 119 | 2017/10/31 | DIS |

The following steps are envisioned for the participation in the call for proposals:

* All proposals shall be prepared in accordance with the requirements provided in Appendix A
* All proposals will be evaluated according to the procedure described in the CDVA Evaluation Framework for Search & Retrieval [2].
* It is expected that proponents produce results by using tools and procedure described in the evaluation framework. Proposals bringing partial results, or results produced in manner that is different from CDVA evaluation procedure, may also be considered and evaluated as part of Core Experiment process.
* In order to participate and get access to the evaluation framework and test material, proponents will be required to register their intent to participate.

**2.2. Subscription to reflector**

Proponents are required to subscribe to CDVS/CDVA AHG reflector

<https://mailhost.tnt.uni-hannover.de/mailman/listinfo/cdvs>

**2.3. Intent to participate**

Interest in participation should be expressed by Monday, 2015/10/12 (24:00 GMT) by sending an email to the CfP Contact (see below). Email should indicate contact names and company (companies).

**2.4. Distribution of test material**

Test items (visual databases) and evaluation software will be available as described in the CDVA Evaluation Framework [2]. Interested parties may also request test materials to be mailed to them. Such requests shall be sent to AHG reflector, the details of the distribution procedure will be provided on request.

The CDVA AHG reflector will be used for communication of any changes in the datasets or annotations used for verification experiments. Requests for changes in annotations must be well justified. Examples of such justifications include:

* offensive content identified;
* true matches detected in the distractor dataset.

**2.5. Submission of proposals**

Proposals shall be submitted by Monday, 2015/02/15 (24:00 GMT) (1 week prior the 114th meeting).

The proposals are expected to include the following material:

1. Technical description of proposal;
2. Completed information forms as provided in Appendix A

(Information on fulfillment of requirements and performance of proposal);

1. Files generated by running experiments as described in Evaluation Framework, including binary descriptor files, indices, and output text files with performance and timing results;
2. Executable software modules used to conduct experiments;
3. Any other relevant information to help the evaluation of the proposal, for example, description of unique features and uses of proposed technology
4. Completed Patent Statement and Licensing Declaration form: <http://www.iso.org/patents>.

Proponents shall submit these materials (except item 3) as input contributions to the 114th MPEG meeting and send title and author information to the CfP Contact prior to the submission deadline.

Software binary modules shall be part of the input contributions. Binary files generated during the experiments (such as descriptor or index files) shall be made available by proponents at the 114th MPEG meeting. Further details will be provided in the evaluation framework.

**2.5. Evaluation of proposals**

Evaluation of proposals will be conducted at the 114th MPEG meeting and during CDVA AHG period before the 115th MPEG meeting. The proposed technologies will be evaluated using the process set forth in the Evaluation Framework document [2]. Proponents are expected to be present personally during the 114th MPEG meeting.

Proponents are advised that MPEG retains the right not to accept any proposal, to accept one or more than one proposal. In any case, MPEG intends to elaborate on the proposals retained in order to produce the Standard.

**3. Source Code and IPR**

Proponents of technologies that are selected for further evaluation will be required to provide source code implementation of all modules needed for reproduction of the submitted results. Such code should include the header of the modified BSD license available at http://mpeg.chiariglione.org/docs/mxm-license. Additionally, submission of improvements (bug fixes, etc.) is certainly encouraged.

Furthermore, proponents are advised that this Call is being made subject to the patent policy of ISO/IEC (see [ISO/IEC Directives Part 1](http://isotc.iso.org/livelink/livelink?func=ll&objId=4230455&objAction=browse&sort=subtype), Appendix I) and the other established policies of the standardization organization.

# 4. Call for Proposal Contact

To register for this Call or for any other questions concerning the Call, contact:

Prof. Dr. Jörn Ostermann

Leibniz Universität Hannover, Institut für Informationsverarbeitung

Appelstr. 9A, 30167 Hannover, Germany

Tel: +49 511 762-5316, Fax: +49 511 762-5333, email [ostermann@tnt.uni-hannover.de](mailto:ostermann@tnt.uni-hannover.de)

Prof. Dr. Jens-Rainer Ohm

RWTH Aachen University, Institute of Communications Engineering

Melatener Str. 23, 52074 Aachen, Germany

Tel. +49-241-8027671, Fax. +49-241-8022196, email [ohm@ient.rwth-aachen.de](mailto:ohm@ient.rwth-aachen.de)

# References

1. N15042, Compact Descriptors for Video Analysis (CDVA), Oct. 2014, Strasbourg, FR
2. N15338, Evaluation Framework for Compact Descriptors for Video Analysis - Search and Retrieval, Jun. 2015, Warsaw, PL
3. N15040, Compact Descriptors for Video Analysis: Requirements for Search Applications, Oct. 2015, Strasbourg, FR.
4. N14507, Compact Descriptors for Video Analysis: Objectives, Applications and Use Cases, Apr. 2014, Valencia, ES.

These documents and Information about MPEG and MPEG standards may be accessed at:

<http://www.chiariglione.org/mpeg/meetings.htm> or directly from ISO at <http://www.iso.org>

# Appendix A. Information Forms to be filled in by the contributors

**Information form**

1. Title of the proposal
2. Organization name
3. Application scenarios addressed by proposal
4. Availability of software modules needed for evaluation of the proposal
5. Information on additional functionality (such as use of side information, etc.) supported by proposal
6. Information on parts of the proposal that must be defined as normative to ensure interoperability.

**Requirements**

| **Requirement** | **Description** | **Fulfilment information** |
| --- | --- | --- |
| **High Matching Accuracy** | Descriptors shall allow matching, retrieval and ranking of video content containing the same objects/scenes as a query image/video. The objects of interest comprise planar or non-planar, rigid or partially rigid, textured or partially textured objects. | Measured in the experiments. |
| **Self-Containment** | Descriptors shall be self-contained, in the sense that no other data are necessary for matching. | Y/N |
| **Coding Independence** | Descriptors shall be independent of the image and video coding formats. | Y/N |
| **Non-Alteration of the Content** | Descriptor extraction and matching shall not require any alteration of the original content or from which the descriptors are extracted or embedding of information into the original content. | Y/N, specify any needed preprocessing |
| **Compactness** | Descriptors shall be compact and shall not exceed a specified average and maximum byte size per frame or group of frames. | Average and maximum descriptor lengths for the dataset. |
| **Size Scalability** | Descriptors shall be scalable and shall be specified at multiple byte sizes per frame/group of frames.  Descriptors of different sizes shall interoperate. | Range of descriptor lengths and database sizes supported. |
| **Low Extraction Complexity** | Descriptor extraction shall be possible with low complexity in terms of memory footprint and computational complexity. | Timing information from descriptor extraction. |
| **Low Matching Complexity** | Descriptor matching shall be possible with low complexity in terms of memory footprint and computational complexity. | Timing information from retrieval and verification experiments. |
| **Spatiotemporal Localization** | Descriptors shall allow localization of matched objects/scenes spatially and temporally in the video stream. | Measured in the experiments. |
| **Object Tracking** | Descriptors shall allow low complexity tracking of matched objects/scenes (e.g. by tracking local descriptors). | Y/N |
| **Partial Temporal Matching** | Descriptors shall support matching of objects/scenes appearing only in a temporal segment of an arbitrarily long query video to the same objects/scenes appearing only in a temporal segment of an arbitrarily long database video. | Measured in the experiments. |
| **Robustness to Image/Video Spatiotemporal Characteristics** | Descriptors shall be robust to differences in source image/video characteristics, i.e. spatial resolution, interlaced/progressive, temporal resolution, spatial aspect ratio. | Measured in the experiment, the modifications considered and their parameters are listed in [2]. |
| **Robustness to Image/Video Capture Conditions** | Descriptors shall support accurate matching in the presence of changes in vantage point, camera parameters, lighting conditions, motion characteristics, film grain, as well as partial occlusions. |
| **Robustness to Image/Video Editing Operations** | Descriptors shall support accurate matching in the presence of editing operations such as compression, transcoding, color changes, and frame rate changes. Text overlay, logos, etc. |