Building video queries for

Trecvid2008 copy detection task

I. Principles

Each query will be constructed by taking a segment of varying length from the test collection, embedding it in a varying amount of video which does not occur in the test collection, and then optionally applying one or more transformations to the entire query segment. Some queries may contain no test segment; others may be composed entirely of the test segment.

We have as raw input materials

- Reference video database (Sound & Vision (200 hrs))
- Videos not in the database (to test false positive)

The video queries can be three types:

- Type 1: the clip is a transformed fragment of reference data
- Type 2: the clip contains a transformed fragment of reference data
- Type 3: the clip is a fragment of a video not in the reference database

This document aims to detail how we want to generate these video queries. In the rest of this document, we use the following notations:

Video number X from the reference database	Ref_X
Video number Y not from the reference database	NonRef_Y
Video query number N with the transformation number T	Query_transT_N

II. Description

The general framework for building video queries is presented in figure 1. The different steps are described in details bellow.
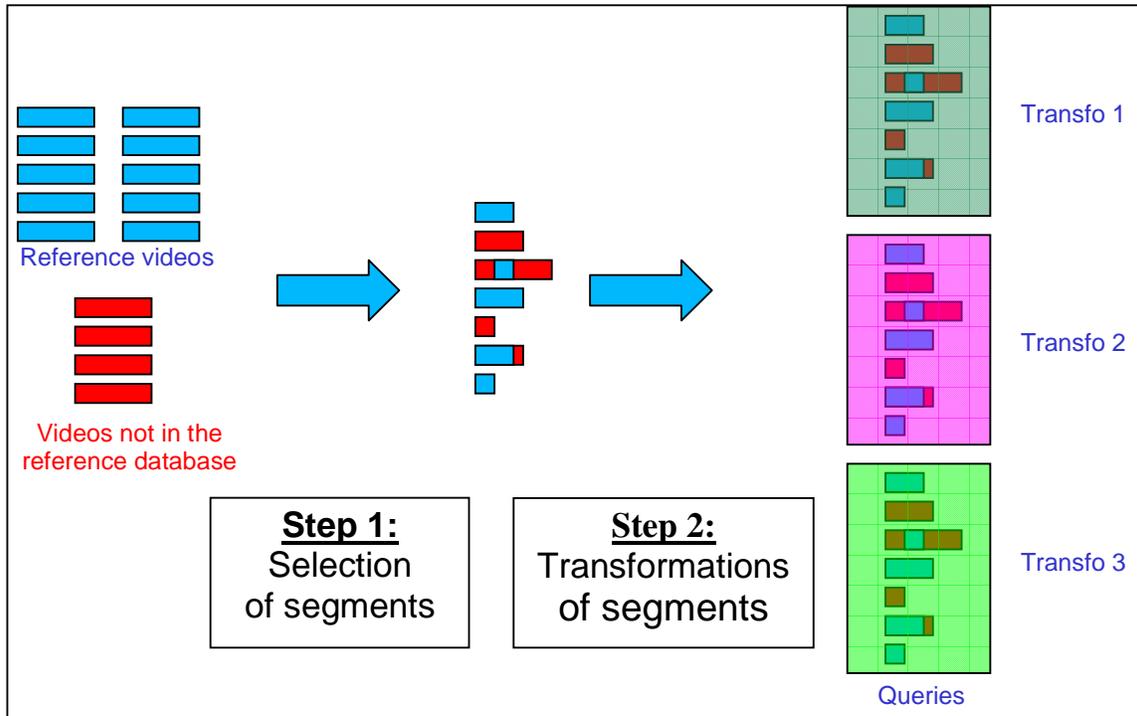
figure 1: framework for building video queries

1. Step 1

Software will be developed at INRIA to achieve step 1, which consists of randomly selecting segments of videos from the two sets (reference video and non reference video) and creating videos with the three types. A ground truth file is associated to each created video. For the construction of the official queries, the software will be used by NIST, which will retain the ground truth for use in the evaluation.

The principle of this software is the following:

- A random Ref_X is selected
    i. A random beginning time code is selected Tc_{Ref}
    ii. A random length is selected L_{ref}

- A random NonRef_Y video is selected
    i. A random beginning time codes is selected Tc_{NonRef}
    ii. A random length is selected $L_{NonRef} > L_{ref}$

- A random type is chosen
    i. If type 1 is selected: we only keep the video segment of Ref_X
    ii. If type 2 is selected, the video segment of Ref_X is inserted in the video segment of NonRef_Y at a random time code.
    iii. If type 1 is selected: we only keep the video segment of NonRef_Y

We propose the following limits for the parameters:

$$3 \text{ seconds} < L_{ref} < 1 \text{ minute}$$
$$30 \text{ seconds} < L_{NonRef} < 3 \text{ minutes}$$

For the choice of Tc_{Ref}, we have to pay attention to the fact that the segments from the program intro/exit material are repeated without change. Therefore the software will automatically ignore the first and last 60 secs of each program when selecting reference segments.

After this step, we have a number N of untransformed videos

2.   Step 2

Step 2 is a selection of transformations which will be applied to the video generated by step 1. Some single transformations are selected and described in section 2.1. Combination of transformation are also considered and described in section 2.2.

2.1 Single transformation

The transformations selected are challenging transformations that occur a lot in real life situation. These transformations are listed in table 1.

Transformation	Parameters, comment
Cam-cording	Have to be manually done, angle of the cam can change
Picture in picture	Scale, position
Insertions of patterns	Can be logo, captions, subtitles
Strong re-encoding	Change of resolution and compression codec
Change of gamma	Gamma factor for each channel

Table1. List of single transformations

Cam-cording: this transformation is done manually by using filming a movie on a screen (same as in the CIVR live benchmark). 3 different angles are considered: camera facing the screen, camera with a small angle and camera with a large angle. These angles will be chosen randomly.

Picture in picture: the untransformed videos queries done by step 1 (from the three types) are inserted in a video not in the database with a change of the scale (reduction); the parameters are the value of this change of scale and the spatial position of the inserted video.

Insertion of patterns: different patterns are inserted randomly: captions, subtitles, logo, sliding captions; a set of logos will be used to be inserted randomly.

Strong re-encoding: the resolution of the video is reduced, the bit rate is changed and we can also encode the video with a different codec; at the end, the video is re-encoded in mpeg1.

Change of gamma: the gamma value for each color is changed randomly

2.2 Combinations of transformation:

Other transformations occur but are less challenging if considered alone. We propose here to do different combinations of transformations. The first kind of combination aims to simulate videos found on the internet with a decrease of quality (transformation are unintentional). The second kind of combination aims to simulate intentional transformation as post production transformation. For these two kinds of combination, we propose to create two sets of queries: one with light transformations (the parameters of each transformation can be chosen in a low range) and another one much more challenging with higher transformation parameters (still randomly selected). Finally we can also create a combination of all the transformations considered.

- Combination 1: decrease in quality

Blur, gamma, color, frame dropping, contrast, change of compression, change of ratio, noise
   - One with light transformations
   - One with hard transformations

- Combination 2: Post production

Crop, shift, change contrast, slow motion, caption, flip (horizontal mirroring),picture in picture.
   - One with light transformations
   - One with hard transformations

- Combination 3 : all the transformations are considered

2.3 Random selection of parameters

In order to have statistics on how the systems deal with the different transformations, we propose to apply each chosen transformation or combination of transformation to all the videos generated in step 1 but with different parameters.

Software developed at INRIA will randomly generate sets of parameters for each of the 10 transformations. The output of the software is a file of the kind:

NonTransformedQuery1
Transfo1: param1 param2 …
Transfo2: param1 param2 …
…
Transfo10: param1 param2 …

NonTransformedQuery2
Transfo1: param1 param2 …
Transfo2: param1 param2 …
…
Transfo10: param1 param2 …

…

NonTransformedQueryN

Transfo1: param1 param2 …
Transfo2: param1 param2 …
…
Transfo10: param1 param2 …

For the combinations of transformations (transfo5 to transfo10), the list of applied transformations is a parameter.

The inputs of the software are the range of the parameters.

For the construction of the official queries, the software will be used by NIST.

3. Step 2

With the parameters of each transformation and combination of transformations for the N query video, independent operator from INRIA will generate the 10 x N video queries. These transformations will be applied as much automatically as possible with probably different software (TMPenc, FinalCut) and manually for the cam-cording. The ground truth might be unknown by the operator if step1 and step2 are achieved by an external partner (e.g. NIST) who will not share the ground truth files.

III. Summary and number of queries

Step 1: 200 videos generated by NIST with the tool developed by INRIA (33% of each of the three types. This number can be changed depending on the chosen metric).

Step 2: File containing the 200 sets of random parameters generated by an external partner with the tool developed by INRIA.

Step 3: A operator of INRIA applies the selected transformations to the 200 videos. At the end, 2000 video queries that will be sent back to NIST

IV. Proposed timeline

April, 15:
Software tools (for step 1 and 2) provided by INRIA to NIST and participants

May, 15:
- Example test queries provided by INRIA to participants
- 200 Non-transformed Video Queries and associated list of transformation parameters provided to INRIA by NIST

June, 31
- Final Video Queries provided by INRIA to NIST