Automatic Video Summarization

Bernard Merialdo , Benoit Huet, Itheri Yahiaoui, Fabrice Souvannavong

Multimedia Communications Department,
Institut EURECOM,
BP 193, 06904 Sophia-Antipolis, FRANCE
{Itheri.Yahiaoui,Bernard.Merialdo,Benoit.Huet,Fabrice.Souvannavong}@eurecom.fr

ABSTRACT[*]

In this paper, we present a new approach to combine text and video in the automatic construction of summaries for audio-video sequences. This extends our previous work which was based on video only. Here, we generalize the Maximum Recollection Principle we were employing and we show how this same principle can be used to build summaries from text, based on several strategies. We present some experimental results in the evaluation of these strategies. Finally, we show how this novel principle can be used to combine text and video information for automatic summarization.

Keywords

Video Analysis, Automatic Summarization, Evaluation

1. INTRODUCTION

Summaries play a useful role in the management of large quantities of documents. For example, they may be used to rapidly find a piece of information (if it is present in the summary), to make a decision (should the complete document be read or not), to evaluate a document (whether it deals with the right topic or not), etc… Automatic summarization is therefore a useful component for information access. Automatic text summarization has been studied for a long time [4][8], yet the results are still imperfect because of the difficulty of Natural Language Understanding. More recently, the automatic summarization of audio-visual sequences has been the subject of increasing research, motivated by the rapid expansion of multimedia information.

While some approaches are taking several media into account (text and video [5], video and audio [7]), much of the research has been devoted to the summarization of multimedia sequences using the video component only [1][3][6][9][12][13]. Several criteria have been proposed to identify important moments in the original sequence, so that the summary can be built from these segments only. Such approaches lead to summaries which are optimal in some mathematical sense, but a common problem, which is still largely unsolved, is to relate these criteria to a user-intuitive measure of quality.

In previous work [9][10][11], we have proposed an approach based on video frame similarity for the automatic creation of video summaries. In this paper, we generalize this approach so that it can be applied to text information, and also to a combination of video and text.

This paper is organized as follows. In the next section, we describe the Maximum Recollection Principle which is used for automatic summarization. In section 3, we briefly expose how our previous work used this principle with video information. Then in section 4, we present how this same principle can be applied to text. This leads to several variants, for which a set of experiments is presented in section 5. In the final section, we describe how the MRP can as well be used with the combination of both video and text, an approach which we are currently investigating.

2. MAXIMUM RECOLLECTION PRINCIPLE

A summary is a subset of the original document, in our case an audio-video sequence. The summary should contain the most "important" information, but the "importance" criterion is very difficult to define. In fact, this criterion depends on the purpose of the summary: the summary should facilitate the user in the realization of a certain task, for example searching some information, selecting a document, making a decision etc… Any subset of the original document is therefore a potential summary, whose quality might be good or bad, depending on the task that is considered (and the careful choice of the subset).

[*] This research was supported by Eurecom's industrial members: Ascom, Cegetel, France Telecom, Hitachi, ST Microelectronics, Motorola, Swisscom, Texas Instruments and Thales.

Our approach is based on a task that we feel relevant to many applications of summaries: the user is asked to identify if a short clip comes from the original audio-video sequence or not, using only the knowledge of the summary (rather than the full sequence). The performance of the user is the percentage of correct decisions over all possible clips taken from the original sequence. We call this task a Maximum Recollection Task (MRT), in the sense that the summary should let the user identify as many clips as possible. The best summary is therefore chosen according to a Maximum Recollection Principle (MRP).

This principle can be formalized as follows:

- Let D be a document (audio-video sequence),
- Let S be a summary (subset) of D,
- Let C be a random clip (continuous subset of D) extracted from D,
- We assume that the user has a decision rule d(C,S) which lets him decide whether a clip C comes from the same document as the summary or not (d=1 for yes , d=0 for no),
- The performance on the MRT is then the average value of d(C,S) over all possible clips C taken from the document D,

$$\text{perf}(S) = \underset{C}{\text{average}}\, d(C,S)$$

With this definition, the best summary according to the MRP is

$$\hat{S} = \underset{S}{\text{argmax}}\, \text{perf}(S)$$

Note that this approach relies on the definition of the decision rule d(C,S). We will provide our choices for video and text-based decisions later. The current presentation generalizes our previous work on video.

The advantage of this approach is that the performance of the summary has a direct intuitive interpretation, and is not only an abstract number. A good summary will allow the user to identify a larger number of clips than a bad summary.

Of course, once the performance criterion has been defined, we also have to design efficient procedures to construct the best summary, or when this is not feasible, a sub-optimal one. This will be described in the application for text summaries.

In practice, we will impose constraints when computing the average and argmax. We will impose the duration of the clips to be identified, and we will impose the duration of the summary (in terms of length or number of keyframes). Those two values are parameters in the automatic summarization procedure.

3. MAXIMUM VIDEO RECOLLECTION

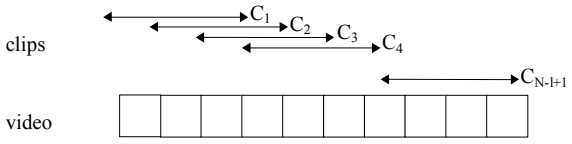
In our previous work [9][10][11], we used this principle on video information only (at the time we called our approach Simulated User, but this relies on the same principle). As mentioned previously, this requires to define the decision rule $d(C^V,S^V)$.

We chose the following rule:

$$d(C^V, S^V) = \begin{cases} 1 & \text{if at least one frame in } C^V \text{ is similar to one frame in } S^V \\ 0 & \text{otherwise} \end{cases}$$

The motivation is that the user can be sure that the video clip C^V comes from the document D when he recognizes an image which is already present in the video summary S^V.

We compute the performance by taking the average over all clips of given duration l, as shown in the following figure:



Similarity between frames was computed using color blob histograms [15]. Additionally, we proposed several methods to build sub-optimal summaries with the aim of maximizing this recollection rate.

Intuitively, the resulting summary contains frames for which similar frames are the most frequent throughout the entire video. The clip constraint helps in lowering the importance of frames which are only "locally" frequent.

Obviously, color blob histograms are a rudimentary measure of image similarity. Through a set of real user experimentations, we could observe that users have a much higher performance rate than the one found automatically, mainly because they are able to interpret the content of the image, such as recognizing people, detecting special clothing or environment, make inferences on sequential events, etc… Nevertheless, the automatic procedure remains a reliable way to select reasonable summaries.

4. MAXIMUM TEXTUAL RECOLLECTION

4.1 Definition

We now describe a new proposal, which is to use the same MRP for text information. We assume that the transcription of the video sequence is available (it could be obtained

through manual transcription, caption extraction or speech recognition) and aligned in time (synchronized) with the video. The summary S^T is now a set of keywords or phrases. A clip C^T is now a short sub-sentence of the original transcription. To apply the MRP, we have to define the decision rule which can be used to decide whether the clip comes from the document or not. We propose the three following variants for this decision rule, representing different policies in evaluating the evidence found in the summary, from weak to strong.

The user decides that this clip is taken from the original document:

- when at least one word of the clip belongs to the summary (**weak policy**),
- when all words of the clip belong to the summary (**strong policy**),
- when at least n words of the clip are present in the summary (**intermediate policy**).

The weak policy is identical in spirit to the rule used for video. However, while it seems perfectly valid for video, because identical images are rare, it is not as obviously correct for text because the occurrence of a given word is not necessarily a firm indication of similarity. This is the reason why we also defined the strong and intermediate policies.

4.2 Summary construction

Having clearly expressed the Maximum textual Recollection Princple, we may now present the textual summary construction methodology in a formal manner. Again, the key factor to the selection of the words of the summary is the decision rule. As a matter of fact, a decision rule is to be defined for each of the above mentioned policies. However, the basic methodology remains the same; the objective is to maximize the performance of the selected words for the summary $\widehat{S}$ as described in section 2.

For textual summarization, we assume that words have been clustered into "similarity" classes W. In our current approach, we consider that two words w_i and w_j are identical if and only if they are identical, except for case:

$$W(w_i) = W(w_j) \quad \text{iif} \quad w_i = w_j$$

(as it is common in Information Retrieval, words that belong to a predefined stop-list of common words are excluded from this process and simply ignored). For the remainder of this paper we will use W_i and w_i interchangeably. In future work, we plan to use more elaborate classification, for example based on stemming, using part-of-speech information etc…

The optimal text summary of size k can be found by enumerating all possible sets of k word classes $\{W_1, W_2, \ldots W_k\}$ and keeping the one which maximizes the average performance over all possible clips C of the document D. Because the enumeration is likely to be computer intensive, it is profitable to select carefully the order in which classes are selected, so that a good solution (sub-optimal) is found early. In practice we have made use of a greedy-like algorithm, in order to iteratively select the summary items. Based on this approach, the performance is decomposed by the following formula:

$$\begin{aligned} \text{perf}(S) &= \text{perf}\left(W_1, W_2, \ldots, W_k\right) \\ &= \text{perf}\left(W_1, W_2, \ldots, W_{k\text{-}1}\right) \\ &\quad + \text{perf}\left(W_k \middle| W_1, \ldots, W_{k\text{-}1}\right) \end{aligned}$$

The sub-optimal summary construction algorithm proceeds as follows:

- Step 1: start with summary S empty,
- Step 2: sort the classes that have not yet been selected by decreasing value of the performance perf(S) with respect to the current summary constitution,
- Step 3: add the class W for which the performance is maximal to the summary. Return to step 2 until summary completion.
- Step 4 (optional): in place refinement of the summary. Take each summary item in turn and attempt to identify a class which improves the performance. This step is repeated until no further improvement can be made.

Note that the algorithm starts by selecting the class W_1 with maximal performance value perf(W_1), then W_2 with maximal performance value perf(W_1,W_2), and so on until W_k. The first complete solution found is the result of a series of greedy choices, and our experiments have shown that it is often an optimal choice over all possible combinations. That is to say that step 4 of the algorithm may be omitted in most cases.

This procedure is fine for the weak policy. However, for the intermediate and strong policies, note that the first word to be selected will have a performance of zero (unless clips are of length 1). So we replace in those cases the exact perf(S) by a proportion of it, linearly depending on the number of words from the clip already in the summary. This allows to select the most promising words, even for the first choice of w_1.

Once the best set of word classes has been found, it only remains to replace each class W_i by its representative word w_i so that the set of words which compose the summary is defined.

As indicated earlier, the computation of the performance function perf(S) depends on the policy used for the creation and the evaluation of the summary. More particularly, it is

the decision rule d(C,S) which is devised specifically for each of the three envisaged policies:

4.2.1 Weak policy:

The idea behind the weak policy is to create a summary for which the largest possible number of clips contains at least one word from the summary. Therefore, it is judicious to select classes with high performance over the original text according to the following decision rule:

$$d_1^T(C^T, S^T) = \begin{cases} 1 & \text{if } \exists w_i \in C^T \text{ s.t. } w_i \in S^T \\ 0 & \text{otherwise} \end{cases}$$

This decision rule encourages little redundancies and high complementarities within the word that make up the summary.

4.2.2 Strong policy:

The strong policy is one where the performance corresponds to the average number of clips for which all word w_i can be found in the summary. The decision rule in this case becomes:

$$d_2^T(C^T, S^T) = \begin{cases} 1 & \text{if } \forall w_i \in C^T \quad w_i \in S^T \\ 0 & \text{otherwise} \end{cases}$$

4.2.3 Intermediate policy:

The intermediate policy states that a good summary should include words such that many clips contain at least n words from selected classes with $1 < n \leq k$. To build a sub optimal summary based on this policy, we propose to devise the decision rule based on the following heuristic:

$$d_3^T(C^T, S^T) = \begin{cases} \dfrac{cov_1}{k} & \text{if } cov_1 \geq n \\ 0 & \text{otherwise} \end{cases}$$

where $cov_1 = card\left(w \in C^T : \exists i \, 1 \leq i \leq k \; w_i \in C^T\right)$

The methods presented up to this point for text summary construction only take into account words from the text under consideration. This may lead to summaries containing frequent words, words which may not be discriminatory enough to "uniquely" represent the original document among others. Although we use a stoplist to remove the most common words, this effect cannot be completely suppressed.

4.3 Contextual summarization

Information Retrieval research has introduced measures such as the Tf.Idf indicator to combine the effect of word frequency and word discriminative power. To transpose it to our problem of Maximal Recollection, we have to consider that, in the decision rule to identify a clip, words which may frequently appear in other documents should be penalized. We can easily adapt our approach to incorporate this new constraint. For example, assume that a personal video library contains several different videos, each one with its corresponding text (transcript or closed caption). To avoid ambiguity, we should include in the summaries of those videos words which are not in the other video transcripts (or at least not in their summaries). In the selection of a word for a specific summary, we should therefore take into account the probability that this word may appear in another summary. When we construct a specific summary, we name as "context" the set of existing summaries in the library.

For the video stream, we have already proposed a methodology to construct multi-episodes video summaries [10][11]. As far as the text is concerned, it is important for the summary to exclude common words, word present in both the current text and the context, but retain specific and content words, which uniquely describe the content of this particular document with respect to others.

In order to estimate the probability that a particular word also appear in the context, we make a very crude assumption by computing the probability that this word appear in r documents randomly selected from the World Wide Web. Such probability can be estimated from indications given by Internet search engines. For the purpose of this experiment we have used the indication provided by goggle (www.google.com) about the frequency of words in the web. Therefore, we compute the probability that the context does not contain the word w_i as:

$$P(w_i \notin context) = (ND - ND(w_i)/ND)^r$$

Where ND is the size of our corpus documents (the part of the Web indexed by Google) and ND(w) is the number of documents in the corpus which involve at least one occurrence of the word w_i (as indicated by Google in the result page after a search on w_i).

We may now re-write the decision rule corresponding to each of the three policies to reflect the effect of the use of contextual information.

4.3.1 Weak policy:

In the case of the weak policy the decision rule is computed as follows:

$$d_4^T(C^T, S^T) = \begin{cases} \max P(w_i) & \text{if } \exists w_i \in C^T \text{ s.t. } w_i \in S^T \\ 0 & \text{otherwise} \end{cases}$$

This means that ambiguous words (which have a low probability of not appearing in the context) will have lower chances of being selected, since they contribute less to the performance of the summary.

4.3.2 Strong policy:

The decision $d_2^T(C^T, S^T)$ is modified in order to account for the addition of contextual information in the following manner:

$$d_5^T(C^T, S^T) = \begin{cases} P(\mathbf{w}) & \text{if } \forall w_i \in C^T \quad w_i \in S^T \\ 0 & \text{otherwise} \end{cases}$$

where

$$P(\mathbf{w}) = \prod_{w_i \in C^T} P(w_i \notin context) = \prod_{w_i \in C^T} (ND - ND(w_i)/ND)^r$$

.

4.3.3 Intermediate policy:

Finally, the probability that words do not belong to the context may be included in the intermediate policy via the following formulation:

$$d_6^T(C^T, S^T) = \begin{cases} \frac{cov_2}{k} P(w_i) & \text{if } cov_2 \geq n \\ 0 & \text{otherwise} \end{cases}$$

where $cov_2 = card(w \in C^T : \exists i \, 1 \leq i \leq k \; w_i \in C^T)$

and $P(w_i)$ is the probability that all the words w_i in the clip C^T are not in the context.

5. EXPERIMENTS

For our experiments, we manually created the transcript of the audio stream of a documentary called "Histoire d'eau" which is part of a video corpus distributed by INA (French National Institute for Audio-Visual). This transcript is composed of 4852 words; as pre-processing all stop words were eliminate which leads to a text containing 1813 words.
We constructed and evaluated summaries of length twenty (each summary contains 20 words) using different clip lengths (from one to eight words) and according to the three policies defined in section 4.

First, we consider the case where no context information is employed.

Table 1: summary performance (construction and evaluation without using context information)

Clip length	Weak Policy d^T_1	Intermediate Policy d^T_3 n>=50%	Intermediate Policy d^T_3 n>=75%	Strong Policy d^T_2
1	16%	16%	16%	16%
2	28.7%	27.6%	4.1%	4.1%
3	40.1%	9.4%	1.6%	1.6%
4	49.4%	13.4%	3.6%	1.1%
5	57.5%	4.5%	1.8%	1.2%
6	64.3%	4.3%	1.7%	1.4%
8	75.2%	2.8%	1.6%	1.2%

Table 1 represents the performance of the two summaries constructed with the methods previously described, in the non-contextual case. Obviously, for the simplest case of clips of length one, all policies are identical. As the clip length increases, the performance of the weak policy increases rapidly, because more and more clips contain at least one word from the summary. On the contrary, the performance of the intermediate and strong policy decreases, since it is more difficult to find clips whose words are all (or in large proportion) in the summary. Note that the fluctuation in performance for the strong policy results are due to the heuristic choice in the greedy procedure, leading to a local optimum. In the intermediate policy experiments, performance variation is also due to the quantization effect for the minimum number of words required.

Table 2: summary performance (construction without context and evaluation with context information)

Clip length	Weak Policy d^T_1	Intermediate Policy d^T_3 n>=50%	Intermediate Policy d^T_3 n>=75%	Strong Policy d^T_2
1	9.80%	9.80%	9.80%	9.80%
2	11.56%	10.60%	2.19%	2.19%
3	10.13%	2.72%	0.52%	0.85%
4	7.86%	2.55%	0.72%	0.21%
5	6.35%	0.65%	0.30%	0.21%
6	4.48%	0.36%	0.18%	0.15%
8	2.33%	0.11%	0.05%	0.05%

In table 2 are shown the performance results of the same summaries as constructed for table 1 when evaluated with a context of ten documents. Those results indicate that the context free construction of text summaries lead to common words being employed since the performance in table 1 are much higher than in table2. In this case, even the performance of the weak policy decreases, since it becomes more and more frequent to find at least one word of the context in the clip as the clip length increases.

Finally, table 3 give the performance results in the case where the summaries are both constructed and evaluated using the contextual measures previously defined. When compared with table 2, those results show that the performance can be slightly improved when summaries are built in a manner which is consistent with the evaluation measure.

Table 3: summary performance (construction and evaluation using contextual information)

Clip length	Weak Policy d^T_4	Intermediate Policy d^T_6 n>=50%	Intermediate Policy d^T_6 n>=75%	Strong Policy d^T_5
1	11.41%	11.41%	11.41%	11.41%
2	12.88%	11.91%	2.72%	2.72%
3	11.16%	3.23%	0.85%	0.85%
4	8.99%	2.81%	0.97%	0.55%
5	6.80%	0.80%	0.60%	0.50%
6	4.98%	0.60%	0.40%	0.30%
8	2.38%	0.24%	0.19%	0.14%

A general remark from this results is that the performance levels that are measured are rather low. Except in the case of the non-contextual weak policy, results rapidly fall around the one percent level, and even sometimes less. Since one would expect summaries to contain a substantial part of the information from the document, those figures are probably too low to be useful. Increasing the size of the summary is obviously one way of getting improved performance, but it is not clear how reasonable a larger size would be. In the case where the size of the summary would be in the hundreds, it is not clear that the experiment using Maximum Recollection would appear valid to a user. Certainly, one direction could be to describe a text as a set of topics, and not simple words, to that the measure is based on the presence of the topics, rather than simple words, in the summary (this would require more elaborate text processing methods, such as Latent Semantic Indexing or other classification techniques, and is out of our current research scope).

Using context to detect and remove ambiguous words seems a sensible thing to do, even if the resulting performance levels are low. Indeed, those results are our first attempt at introducing text information in the summarization process, and we plan to expand these experiments further. In particular, the weak policy with context seems more appropriate than the weak policy without context, despite the small evaluation difference. This is in accordance to our expectation since words contained in the summary using context are specific to the related document whereas the summary constructed without context carries only the most frequent (common) words. It is interesting to note that a similar result was obtained while constructing video-based summaries [9].

6. MAXIMUM RECOLLECTION BY VIDEO AND TEXT

The same approach can be further expanded if we consider that we can present the user with both video and text information together.

A video document is a combination of different simultaneous streams (video, audio and eventually closed caption). If we take a random clip C of the video, we can consider both its video component C^V and its textual transcription C^T . The application of our general principle is now to construct an optimal summary S, composed a video part S^V and a text part S^T as depicted in figure 1 (this corresponds to the concatenation of video and text summaries constructed independently). It is clear that much semantic information is gained through the addition of keywords to the set of key frames. The video and text parts are not necessarily synchronized but should be complementary in order to maximize the amount of information from the original video covered by the summary:

$$\hat{S} = \arg\max_{S} \text{ perf } (S^V, S^T)$$

eau, robinet, ville, potable, minérale, Moscou, tuyaux, qualité, Rouslane, New-York, chlore, parasite, boire, eaux, filtrée, fer, canalisations, kilomètres, Kennedy, parisiens

Figure 1: An example of multimedia summary using both textual and visual information

In the combined video-text case, it seems reasonable to define a weak policy such as:

The clip C is correctly identified (as originating from document D) using the summary S, if

- either the video clip C^V is guessed correctly using the video part of the summary S_V
- or the text clip C^T is guessed correctly using the textual part of the summary S^T.

Our future work will be to evaluate this approach, and in particular, to evaluate how different are summaries build with independent video and text criteria, or using a combined video-text criterion. The latter approach should provide better performance, since the summarization process should be able to detect redundancy between video and text and take this into account in the selection of relevant keyframes and keywords.

The same approach can also be used to evaluate video skims, where the text is the exact transcription of the audio track, because the selection of an audio-video segment to be added to the summary can be done based on a combined video-text evaluation measure. We have not yet studied such combination and processes, but we believe that this is a promising approach.

7. CONCLUSION

In this paper, we have proposed a novel approach to automate the creation of multimedia summaries based on the Maximal Recollection Principle. This principle corresponds to an identification task a short clip from the original media using video and/or text summary information. According to this specific task (which is a clearly defined application), we can apply this principle to either the automatic creation and evaluation of a video or a textual summary. Additionally, we introduce the idea of using a document context to build more discriminating text summaries. Our experimental results for textual summarization, show that among the different algorithmic alternatives (the various policies devised and the use or not of contextual information) the most interesting performance is obtained using a weak policy combined with contextual information. Finally, we outline how this same principle can be extended to combine both video and text information simultaneously, which leads to promising new methods for the construction of multimedia summaries.

8. REFERENCES

[1] Anastasios D. Doulamis, Nikolaos D. Doulamis and Stefanos D. Kollias. Efficient video summarization based on a fuzzy video content representation. IEEE International Symposium on Circuits and Systems, Vol. 4, pp. 301-304 May 28-31, 2000.

[2] Bernard Merialdo, Kyung Tak Lee, Dario Luparello, and Jeremie Roudaire. Automatic construction of personalized TV news programs. In ACM Multimedia conference, November 1999.

[3] Emile Sahouria and Avideh Zakhor. Content Analysis of Video Using Principal Components. IEEE Transactions on circuits and systems for Video technology, Vol 9, No 8, pp. 1290-1298, December 1999.

[4] Inderjeet Mani and Mark T. Maybury. Advances in Automatic Text Summarization. The MIT Press, 1999.

[5] Smith M.A. and T. Kanade. Video skimming and characterization through the combination of image and language understanding. IEEE International Workshop on Content-Based Access of Image and Video Database, pp. 61-70, 1998.

[6] Nuno Vasconcelos and Andrew Lippman. Bayesian modeling of video editing and structure: Semantic features for video summarisation and browsing. IEEE Intl. Conf. on Image Processing, Vol. 3, pp. 153-157, 1998.

[7] Rainer Lienhart, Silvia Pfeiffer and Wolfgang Effelsberg. Video abstracting. In Communications of ACM, December 1997.

[8] Udo Hahn and Indejeet Mani. The challenges of automatic Summarization. IEEE Computer, Vol. 33(11), pp. 29-36, November 2000.

[9] Itheri.Yahiaoui, Bernard Merialdo and Benoit Huet. Comparison of Multi_Episode Video Summarisation Algorithms. Workshop on MultiMedia Signal Processing, pp 461-466, 2001.

[10] Itheri Yahiaoui, Bernard Merialdo, Benoit Huet. Generating Summaries of Multi-Episodes Vidéo. IEEE International Conference on Multidedia and Expo, 2001.

[11] Benoit Huet, Itheri Yahiaoui and Bernard Merialdo. Multi-Episodes Video Summaries. International Conference on Media Futures, pp. 231- 234, 2001.

[12] Sundaram, H and Shifu Chang, Constrained Utility Maximization for generating Visual Skims. IEEE Workshop on Content-based Access of Image and Video Libraries. pp. 124-131, Dec 2001

[13] Yihong Gong and Xin Liu. Generating optimal video summaries. IEEE International Conference on Multimedia and Expo, Vol. 3, pp. 1559 -1562, 2000.

[14] Camedir Toklu, Shih-Ping Liou and Madirakshi Das. Videoabstract: A Hybrid Approach to generate semantically Meaningful Video summaries, IEEE International Conference on Multimedia and Expo, Vol 3, pp. 1333-1336, 2000.

[15] Itheri Yahiaoui, Bernard Merialdo and Benoit Huet, Image Similarity for Automatic Video Summarization. To appear in Proc. XI European Signal Processing Conference, Toulouse, France, September 3-6, 2002.