

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 April 2011 (07.04.2011)

(10) International Publication Number
WO 2011/039773 A2

PCT

(51) International Patent Classification:
H04N 7/173 (2011.01)

(21) International Application Number:
PCT/IN2010/000617

(22) International Filing Date:
14 September 2010 (14.09.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2092/MUM/2009 14 September 2009 (14.09.2009) IN

(71) Applicant (for all designated States except US): **TATA CONSULTANCY SERVICES LTD.** [IN/IN]; Nirmal Building, 9th floor, Nariman Point, Mumbai 400 021, Maharashtra (IN).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HIRANMAY, Ghosh** [IN/IN]; TCS Innovation Labs Delhi, TCS Towers, 249 D & E Udyog Vihar Phase IV, Gurgaon 122 015, Haryana (IN). **SUNILKUMAR, Kopparapu** [IN/IN]; TCS Innovation Labs Mumbai, Yantra Park, Pokhran Road No.2, Subhash Nagar, Thane (W) 400 601, Maharashtra (IN).

(74) Agent: **DEWAN, Mohan**; R. K. Dewan & Company, Trade Mark & Patent Attorneys, Podar Chambers, S.A.Brelvi Road, Fort, Mumbai 400 001, Maharashtra (IN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to the identity of the inventor (Rule 4.17(i))
- of inventorship (Rule 4.17(iv))

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: TV NEWS ANALYSIS SYSTEM FOR MULTILINGUAL BROADCAST CHANNELS

(57) Abstract: A system for identification, classification, storage, and analysis of news programs, containing an audio channel, video channel, and metadata relating to it, broadcasted/relayed on a television (TV) channel by means of a plurality of TV broadcast streams, said system comprising: - acquisition module adapted to capture said TV broadcast streams; - recording module adapted to record said captured streams on a physical storage; - news program identification module adapted to identify news programs in said stored broadcast streams; - news program clipping module adapted to separate said identified news programs from other programs; - advertisement identification module for identification of advertisements from said identified news programs; - advertisement clipping module adapted for removal of said identified advertisements; - seam detection module adapted to detect and identify seams of said news programs in order to demark individual stories in a news program; - keyword generation module adapted to generate a list of keywords; - text-keyword identification module adapted to identify said created keywords from visual text of identified said news programs; - speech-keyword identification module adapted to identify the created keywords from the speech of said identified news programs; - repeat-identification module adapted to identify similar/repeat news programs from said plurality of TV broadcast streams; - clustering module adapted to cluster said repeat news programs into one news programs, in order to avoid duplication or multiplication; - removal module adapted to remove said repeat-identified news programs; - repository adapted to store said news programs and metadata embedded in said news programs; and - logical interconnection module adapted to logically interconnecting each of said modules for determining the sequence of steps for a multilingual news video analysis system.



WO 2011/039773 A2

TV NEWS ANALYSIS SYSTEM FOR MULTILINGUAL BROADCAST CHANNELS

Field of the invention:

The present invention relates to the field of computer vision and audio processing techniques.

Particularly, the present invention relates to analysis of television (TV) news channels.

Still particularly, this invention relates to TV news analysis system for multilingual broadcast channels

Background of the Invention:

Round-the-clock monitoring of several news channels in different languages, which is of paramount importance to several agencies, requires unaffordable manpower with language skills and is error-prone because of possible distractions. Thus, it is necessary to have an automated system for identifying and indexing individual stories from news broadcast streams. It is also necessary to filter out repeat transmissions and cluster similar stories broadcast on different channels, possibly in different languages.

While there has been significant research in multimodal analysis of news-video for their automated indexing and classification, the commercial applications are yet to mature. The commercial products like BBN Broadcast

monitoring system and Nexedia rich media solution offer speech-only solutions for TV news broadcast indexing and retrieval for English and a handful of other languages. None of these solutions can differentiate between TV news and other TV programs and additionally they cannot filter out commercials. They index the audio-stream as a whole and cannot demarcate news story boundaries. None of the available solutions can handle telecasts in multiple languages using audio-visual features.

PRIOR ART

The paper, “*Transcribing broadcast news for audio and video indexing*”; Communications of the ACM (CACM), 43(2), Feb 2000. pp 64—70; by Jean-Luc Gauvain, Lori Lamel and Gilles Adda, and another paper, “*Speech and Language Technologies for Audio Indexing and Retrieval*”; Proceedings of the IEEE, 88(8), August 2000 by John Makhoul, Francis Kubala, Timothy Leek, Daben Liu, Long Nguyen, Richard Schwartz and Amit Srivastava propose audio-based approaches, where the speech, in multiple languages, is transcribed and the constituent words and phrases have been used to index the contents of a broadcast stream. This approach has been followed in the commercial systems like BBN and Nexedia.

Another paper, “*TRECVID 2004 search and feature extraction tasks by NUS PRIS*”; NIST TRECVID-2004, Nov 2004; by Tat-Seng Chua, S.Y. Neo, K. Li, G.H. Wang, R. Shi. M Zhao, H. Xu S. Gao and T.L proposes use of

additional sources of information, e.g. Video OCR information, face recognizer and speaker identification for indexing news videos in English.

The paper, “*Detection of acoustic patterns in broadcast news using neural networks*”; Acustica 2004; by H. Meinedo and J. Neto proposes a method for using jingles to mark the boundaries of different programs on a TV channel.

The papers, “*Automatic TV advertisement detection from MPEG bitstream*”; Pattern Recognition, 35(12), December 2002, pp 2719 — 2726; by David A. Sadlier, Sean Marlow, Noel O'Connor and Noel Murphy; “*Time-constraint boost for TV commercial detection*”; International Conference on Image Processing, (ICIP '04),_24-27 Oct. 2004. Vol 3, pp: 1617 – 1620; by Tie-Yan Liu, Tao Qin and Hong-Jiang Zhang; “*Robust learning-based TV commercial detection*”; Proc. 14th ACM International Conference on Multimedia and Expo (ICME), Amsterdam, 6 Jul, 2005; by Xian-Sheng Hua, Lie Lu and Hong-Jiang Zhang; “*Segmentation Categorization and identification of commercials from TV streams using multimodal analysis*”; International Multimedia Conference (MM'06), 23-27 October, 2006; by Ling-Yu Duan, Jinqiao Wang, Yantao Zheng, Hanqing Lu and Jesse S. Jin; “*TV ad video categorization with probabilistic latent concept learning*”; Multimedia Information Retrieval (MIR'07), Augsburg, Sept 2007, pp 217—226; by Jinqiao Wang, Ling-Yu Duan, Lei Xu, Yantao Zheng, Jesse S. Jin, Hanqing Lu and Changsheng Xu propose different methods for detecting advertisement breaks and advertisement classification.

Another paper, “*Story boundary detection in large broadcast news video archives: techniques experience and trends*”; 12th ACM International Conference on Multimedia (MM’04), pp. 656 – 659, 2004; by Tat-Seng Chua, Shih-Fu Chang, Lekha Chaisorn and Winston Hsu surveys several methods for story-boundary identification in a news program. Different methods are further described in “*Story segmentation of broadcast news in English, Mandarin and Arabic*”; Proc. Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, 4-9 June 2006; by Andrew Rosenberg and Julia Hirschberg; “*Story segmentation of broadcast news in Arabic, Chinese and English using multi-window features*”; Proc 30th annual international ACM SIGIR conference on research and development in information retrieval (poster). pp 703 – 704, 2007; by Martin Franz and Jian-Ming Xu; “*Texttiling: segmenting text into multi-paragraph subtopic passages*”; Computational Linguistics, 23(1), pp 33—64. 1997; by M.A. Hearst; “*Unsupervised video-shot segmentation and model-free anchor-person detection for news video parsing*”; IEEE Trans. Circuits and Systems for Video Technology. 12(9), pp. 765 – 776, 2002; by X. Gao and X. Tang; “*Combining text and audio-visual features in video indexing*”; Proceedings of IEEE International Conference on Accoustics, Speech and Signal Processing (ICASSP ’05) pp. 1005—1008, 2005; by Shih-Fu Chang, R. Manmatha and Tat-Seng Chua; “*A multi-modal approach to story segmentation for news video*”; World Wide Web: Internet and Web Information Systems. Vol 6. pp 187—208. 2003; by Lekha Chaisorn, Tat-Send Chua and Chin-Hui Lee; “*Video story segmentation with multi-modal*

features: experiments on TRECvid 2003"; Multimedia Information Retrieval (MIR'04). October 15-16, 2004; by Laurent Besacier, George Quénot, Stéphane Ayache and Daniel Moraru.

Papers, "*A probabilistic framework for TV-news story detection and classification*"; IEEE International Conference of Multimedia and Expo (iCME'05), pp 1350—1355. July 2005; by Francesco Colace, Pasquale Foggia and Gennaro Percannella and "*Using Multimedia Ontology for generating conceptual annotations and hyperlinks in video collections*"; International conference on Web Intelligence, Hong Kong, December 2006; by Gaurav Harit, Santanu Chaudhury and Hiranmay Ghosh, concentrates on sports news classification.

US2004189873 discloses a VIDEO DETECTION AND INSERTION SYSTEM. US2004189873 system includes means which detects defined segments in a video stream. The defined segments may be advertisements, as mentioned.

US6614987 discloses a TELEVISION PROGRAM RECORDING WITH USER PREFERENCE DETERMINATION. The system includes a module which is responsive to attribute information in accordance with categorization (classification) parameters or viewing preferences of the user. Thus, news stories may be identified according to this document.

US2002162118 discloses an EFFICIENT INTERACTIVE TV. This system includes content identifier means to identify content or a subset of content. This identification, not only helps in identify news stories and advertisements, but is also poised to identify repetitive news stories.

US6608930 discloses a METHOD AND SYSTEM FOR ANALYZING VIDEO CONTENT USING DETECTED TEXT IN VIDEO FRAMES. This system detects video streams based on user-selected image text attributes. A selected attribute may be news stories or advertisements or both. So, both may be individually identified for segregation purposes. Further, recognition of persons featuring in the detected video is also disclosed. However, all these features are enabled due to the (image) text that is available in each video stream.

However, none of the above patents / patent applications provide a solution for handling multilingual applications such as multilingual television channels including multilingual news programs and removal of advertisements, thereof.

US2006136226 discloses a SYSTEM AND METHOD FOR CREATING ARTIFICIAL TV NEWS PROGRAMS. It includes means to process the language of a newscaster to be translated into choice of user. combines automatic speech recognition (Speech-To-Text processing), automatic machine translation, and audio-visual Text-To-Speech (TTS) synthesis techniques for automatically personalizing TV news programs. However,

this patent application does not provide a solution for identifying news programs and classification of said programs, or even identifying programs based on metadata.

While there are several methods for different aspects of news video analysis, there is a need for a process to combine these tools for creating a news video analysis solution that obviates the limitations of the prior art.

OBJECTS OF THIS INVENTION

An object of the invention is to provide an integrated and complete solution for news video analysis.

Another object of the invention is to provide a system wherein TV newscasts in different languages can be processed.

Yet another object of the invention is to automatically identify news programs in a broadcast stream and separate it out from other programs.

Still another object of the invention is to provide a system wherein advertisements in TV newscasts are automatically identified and removed from the news program.

An additional object of the invention is to provide a system for new analysis wherein the story boundaries are automatically identified and the news stories are segregated.

Yet an additional object of the invention is to provide a system wherein repeat telecasts are identified and filtered out.

Still an additional object of the invention is to provide a system for news analysis wherein similar stories (pertaining to the same event) on different channels are identified and clustered.

Another additional object of the invention is to provide a system for news analysis wherein each news story is indexed with keywords identified in the speech and visual text as well as other metadata, such as a recognized face.

Another object of the invention is to provide a system where news stories in languages, for which speech and OCR technologies are not mature, are indexed based on their similarity with stories in other languages where speech and OCR technologies is mature.

Yet another additional object of the invention is to provide a system for news analysis wherein the stories are classified and can be retrieved.

SUMMARY OF THE INVENTION

According to this invention, there is provided a system for analysis of news channels/stories broadcasted/relayed on a television (TV).

According to this invention, there is provided a system for identification, classification, storage, and analysis of news programs, containing an audio channel, video channel, and metadata relating to it, broadcasted/relayed on a television (TV) channel by means of a plurality of TV broadcast streams, said system comprising:

- acquisition module adapted to capture said TV broadcast streams;
- recording module adapted to record said captured streams on a physical storage;
- news program identification module adapted to identify news programs in said stored broadcast streams;
- news program clipping module adapted to separate said identified news programs from other programs;
- advertisement identification module for identification of advertisements from said identified news programs;
- advertisement clipping module adapted for removal of said identified advertisements;
- seam detection module adapted to detect and identify seams of said news programs in order to demarcate individual stories in a news program;
- keyword generation module adapted to generate a list of keywords;

- text-keyword identification module adapted to identify said created keywords from visual text of identified said news programs;
- speech-keyword identification module adapted to identify the created keywords from the speech of said identified news programs;
- repeat-identification module adapted to identify similar/repeat news programs from said plurality of TV broadcast streams;
- clustering module adapted to cluster said repeat news programs into one news programs, in order to avoid duplication or multiplication;
- removal module adapted to remove said repeat-identified news programs;
- repository adapted to store said news programs and metadata embedded in said news programs; and
- logical interconnection module adapted to logically interconnecting each of said modules for determining the sequence of steps for a multilingual news video analysis system.

Typically, said keyword generation module is a multilingual keyword generation module adapted to generate keywords in multiple languages.

Typically, said text-keyword identification module is a multilingual text-keyword identification module adapted to identify said created keywords from visual text of identified said news programs, in different languages, in the visual channel of the news program.

Typically, said speech-keyword identification module is a multilingual speech-keyword identification module adapted to identify said created keywords from the speech in different languages, in the audio channel of the news program.

Typically, said system includes a multilingual lexicon database for generating multilingual synonymous keywords for said created keywords.

In accordance with an embodiment of this invention, there is provided an **acquisition module** adapted to capture a TV broadcast stream and further includes a **recording module** adapted to record said captured stream on a physical storage, typically on disk, in chunks of manageable size.

In accordance with another embodiment of this invention, there is provided a **news program identification module** adapted to identify news programs in the broadcast stream and to separate them from other programs.

In accordance with yet another embodiment of this invention, there is provided an **advertisement identification module** for identification of advertisements from said news programs, and further including an **advertisement clipping module** adapted for removal of said identified advertisement breaks.

In accordance with still another embodiment of this invention, there is provided a **keyword generation module** to create a list of desired keywords of contemporary interest in different languages.

In accordance with an additional embodiment of this invention, there is provided a **text-keyword identification module** adapted to identify the desired created keywords from the visual text of said news stories, in different languages, typically appearing in form of ticker text on the screen.

In accordance with yet an additional embodiment of this invention, there is provided a **speech-keyword identification module** adapted to identify the desired keywords from the speech, in different languages, in the audio channel of the news.

In accordance with still an additional embodiment of this invention, there is provided a **seam detection module** adapted to detect and identify seams i.e. story boundaries and demarcate the individual stories in a news program.

In accordance with another additional embodiment of this invention, there is provided a **repeat-identification module** adapted to identify similar/repeat stories from multiple channels and further including a **clustering module** adapted to cluster said repeat stories into one story, to avoid duplication or multiplication.

In accordance with yet another additional embodiment of this invention, there is provided a **removal module** adapted to identify duplicate stories (repeat telecasts) and remove them from the selected stories.

In accordance with still another additional embodiment of this invention, there is provided a **repository** adapted for storing the news contents, content description of the news videos, various indexes and links as discovered in the previously described modules.

Typically, said system includes a retrieval module adapted for retrieving a news program from said repository.

Typically, said system includes a navigation means adapted for navigation in said repository for retrieving a news program.

In accordance with yet another embodiment of this invention, there is provided a **logical interconnection module** adapted to logically interconnect all the said modules for determining the sequence of steps for a multilingual news video analysis system.

Typically, repeat-identification module includes visual matching means adapted to use visual cues in order to identify repeat news programs, said visual matching means comprises:

- key frame identification means adapted to identify at least a key frame in a plurality of news program;
- key frame visual feature extruding means adapted to extrude visual features relating to pre-defined parameters of said identified key frames;
- processing means adapted to process said extruded features based on pre-defined tests in order to obtain a processed similarity score in relation to plurality of news program;
- identification means adapted to identify repeat news programs based on pre-defined criteria of said similarity score; and
- deletion means adapted to delete said identified repeat news programs.

Typically, repeat-identification module included audio matching means adapted to use audio cues in order to identify repeat news programs, said audio matching means comprises:

- window determination means adapted to determine a window of frames in a plurality of news programs;
- fingerprint detection means adapted to detect audio fingerprint based on pre-defined processing criteria on said determined window of frames;

- processing means adapted to process said detected audio fingerprint based on pre-defined tests in order to obtain a processed similarity score in relation to plurality of news program;
- identification means adapted to identify repeat news programs based on pre-defined criteria of said similarity score; and
- deletion means adapted to delete said identified repeat news programs.

According to this invention, there is provided a method for identification, classification, storage, and analysis of news programs, containing an audio channel, video channel, and metadata relating to it, broadcasted/relayed on a television (TV) channel by means of a plurality of TV broadcast streams, said method comprises the steps of:

- capturing said TV broadcast streams;
- recording said captured streams on a physical storage;
- identifying news programs in said stored broadcast streams;
- separating said identified news programs from other programs;
- identifying advertisements from said identified news programs;
- clipping said identified advertisements;
- detecting and identifying seams of said news programs in order to demarcate individual stories in a news program;

- generating a list of keywords;
- identifying said created keywords from visual text of identified said news programs;
- identifying the created keywords from the speech of said identified news programs;
- identifying similar/repeat news programs from said plurality of TV broadcast streams;
- clustering said repeat news programs into one news programs, in order to avoid duplication or multiplication;
- removing said repeat-identified news programs;
- storing said news programs and metadata embedded in said news programs; and
- logically interconnecting each of said modules for determining the sequence of steps for a multilingual news video analysis system.

Typically, said step of identifying said created keywords from visual text of identified said news programs includes the step of generating keywords in multiple languages.

Typically, said step of identifying created text keywords includes the step of identifying created keywords from visual text of identified said news programs, in different languages, in the visual channel of the news program.

Typically, said step of identifying speech-keywords includes the step of identifying said created keywords from the speech in different languages, in the audio channel of the news program.

Typically, said method includes a step of retrieving a news program from said repository.

Typically, said method includes a step of navigating in said repository for retrieving a news program.

Typically, said step of removing repeat-identified news programs includes a method of using visual cues in order to identify repeat news programs, said method comprises the steps of:

- identifying at least a key frame in a plurality of news program;
- extruding visual features relating to pre-defined parameters of said identified key frames;

- processing said extruded features based on pre-defined tests in order to obtain a processed similarity score in relation to plurality of news program;
- identifying repeat news programs based on pre-defined criteria of said similarity score; and
- deleting said identified repeat news programs.

Typically, said step of removing repeat-identified news programs includes a method of using audio cues in order to identify repeat news programs, said method comprises the steps of:

- determining a window of frames in a plurality of news programs;
- detecting audio fingerprint based on pre-defined processing criteria on said determined window of frames;
- processing said detected audio fingerprint based on pre-defined tests in order to obtain a processed similarity score in relation to plurality of news program;
- identifying repeat news programs based on pre-defined criteria of said similarity score; and
- deleting said identified repeat news programs.

BRIEF DESCRIPTION OF THE ACCOMPANYING DRAWING

The invention will now be described in relation to the accompanying drawings, in which:

Figure 1 illustrates a schematic block diagram of the multilingual news video analysis system.

DETAILED DESCRIPTION OF THE INVENTION

Figure 1 illustrates a schematic block diagram of the multilingual news video analysis system in accordance with the present invention.

The **Telecast Acquisition Module (10)** captures telecast from several possible sources, e.g. a DTH dish, cable TV, etc., tunes to a particular channel, decodes the TV signals and converts the transmission in standard digital video format, e.g. MPEG-4 or the like. This module is replicated for every channel to be monitored.

The video streams captured by the **Telecast Acquisition modules (10)** are stored in a **Recording Module (20)** in chunks of manageable size with unique file names.

These video chunks are then fed to a **News Program Identification Module (30)**, which has pre-fed knowledge about schedule of news programs and jingles that precede and follow the news programs in the various channels, and the like. The module marks the beginning and end of the news programs. This information is stored, typically in standard MPEG-7 compliant format, in **Video Description Module (40)**.

Advertisement breaks within a news program are now detected using absence of specific ticker-text bands and marked in the video in **Advertisement Identification Module (50)**. At this stage, the video is

decomposed into constituent shots and several visual and audio parameters are extracted. The additional information accumulates in **Video Description Module (40)**.

A set of keywords of contemporary interest are selected by analysis of RSS feeds by a **Keyword Generation Module (60)**. The video segments representing news programs are now processed to detect these keywords. A **Keyword Recognition Module (70)** analyzes the visual text and speech to spot the identified keywords. The visual keywords are classified into 'global' and 'local' categories, depending on the ticker-text band where they appear. While the 'local' keywords pertain to the current story being telecast, the 'global' keywords do not pertain to a story that may appear anywhere in the news program. **Speaker identification Module (80)** identifies the speaker using face recognition and speaker identification (speech) technologies in the scenes containing one dominant speaker, for example in speeches made by important personalities. The additional information further augments the description in **Video Description Module (40)**.

Typically, the keyword generation module is a multilingual keyword generation module.

The system provided an ability process multilingual news programs, according to this invention. A multilingual keyword list, in multiple languages, is created, in order to enable keyword spotting in multilingual TV news broadcast channels, both in spoken and visual forms. The

```

<RULE NAME="KeyWord">
  <L PROPNAME="keyword">
    <CONCEPT NAME="Afghanistan">
      <ENG KEY="Afghanistan">Afghanistan</ENG>
      <BEN KEY="Afghanistan">آفغانستان</BEN>
      <HIN KEY="Afghanistan">अफगानिस्तान</HIN>
      <TEL KEY="Afghanistan">అఫగానిస్తాన్</TEL>
    </CONCEPT>
    <CONCEPT NAME="Rajshekhhar">
      <ENG KEY="Rajshekhhar">Rajshekhhar</ENG>
      <BEN KEY="Rajshekhhar">రాజశేఖర్</BEN>
      <HIN KEY="Rajshekhhar">राजशेखर</HIN>
      <TEL KEY="Rajshekhhar">రాజశేఖర్</TEL>
    </CONCEPT>
    <CONCEPT NAME="Terrorist">
      <ENG KEY="Terrorist">Terrorist</ENG>
      <BEN KEY="Santrasbaadi">సన్త్రాసబాదీ</BEN>
      <HIN KEY="Atankabaadi">आतंकवादी</HIN>
      <TEL KEY="Atankavaadi">అతన్కవాది</TEL>
    </CONCEPT>
  </L>
</RULE NAME>

```

multilingual keyword list helps to automatically map the spotted keywords in different languages to a primary language (say English) equivalents for uniform indexing across multiple channels. Restricting the keyword list to a small number helps in improving the accuracy of the system, especially for keyword spotting in speech. A sample multilingual keyword list is shown below:

The method for creating a multilingual keyword list is fueled by RSS feeds, maintained by some website systems. Typically, RSS feeds captures the contemporary news in a semi-structured XML format and contains hyperlinks to the full-text news stories usually in English. The system of this invention identifies the common (statistical language processing) and proper nouns (using named entity detection processing) in the RSS feed text and the associated stories as the keywords. The keywords in the language of the RSS (usually English) forms a set of concepts, which need to be identified in the audio-visual broadcast in different language telecasts. The equivalent keywords in other languages from the English keywords, can be derived using a word level English-to-language dictionary (for common noun keywords) that language; a pronunciation lexicon (a lexicon is an association of words and their phonetic transcription. It is a special kind of dictionary that maps a word to all the possible phonemic representations of the word.) for transliterating proper names in a semi-automatic matter as suggested. Finally, the keywords in multilingual form is dynamic keyword list structure in XML format. This becomes an active keyword list for the news video channels and is used for both keyword spotting in audio-visual new telecast.

One of the novelties of this invention is the use of keyword spotting instead of adopting a full transcription of new telecast to annotate multilingual news broadcast. This serves three purposes (a) one need not determine the language of telecast a priori and (b) one need not have language specific speech recognition engines and (c) it is easier to keyword spot than try a full

text transcription because the search space of the speech to text (speech recognition engine) is constrained in search space. Additionally, it is sufficient to annotate the news telecast by the keywords because news broadcast is all about places and people (proper nouns) and a set of commonly nouns; additionally keyword annotation of the news broadcast occupies much less space than the erroneous full text transcription.

The **Seam Detection Module (90)** uses the video descriptors available in **Video Description Module (40)** to identify story boundaries. **Repeat-Identification Module (100)** identifies similar and duplicate stories from multiple channels. The OCR and speech technology for many Indian languages are not mature enough for reliable keyword extraction. Similar shot detection helps in classification of news stories in these languages. The additional information further augments the description in **Video Description Module (40)** and is used to create a **Repository Knowledge Base (110)**. The knowledge base enables semantic search for news clusters by semantic analysis of the various metadata associated with the news videos in the earlier stages of processing.

For identifying similar and duplicate news programs both audio and visual cues are identified and used, from a plurality of news programs. Firstly, the recorded news videos are segregated into news programs for further processing. As a first step, shot detection technique is used where the news

stories are logically segmented into distinct shots wherein each shot is represented by a *key frame* or *representative frame*. Further, similar story detection module finds similarity score, using visual matching techniques, between two news stories in the range of $[0, 1]$, where '0' means no match and '1' means complete match. After this, the duplicate story detection module finds whether two news stories are duplicates of each other.

The shots are detected on the basis of difference in visual features of the successive frames in a video. A key frame or representative frame and its corresponding visual features such as colour, texture, edges, etc are extracted for each shot. The shots are clustered based on the visual similarity of their representative frames. This is calculated by distance measures such as Absolute Image Difference, Histogram Intersection, Hausdorff Distance, Color Moments, SIFT, and the like.

Each cluster in a story is now compared with every cluster in the other story by comparing the central representative frames in the clusters using a visual comparator. Let $\{c_{11}, c_{12} \dots c_{1m}\}$ be the clusters in story s_1 , and $\{c_{21}, c_{22} \dots c_{2n}\}$ be the clusters in story s_2 . Let k_{ij} be the number of shots in j^{th} cluster of story i . The process is repeated with every pair of candidate similar stories and clusters of similar stories are discovered. Let $\psi(c_{1i}, c_{2j}) = \{0 \mid 1\}$ represent the match between cluster pair c_{1i} and c_{2j} .

Similarity is defined as:

$$SIM_{12} = \frac{\sum \{(k_{1i} + k_{2j}) * \psi(c_{1i}, c_{2j})\}}{\sum k_{ij}}$$

If SIM_{12} is greater than a certain threshold, the news programs are designated to be similar.

Duplicate stories are a subset of similar stories. Two stories are said to be duplicates of each other only if their audio-visual patterns are same.

Let T_{s1} and T_{s2} be the total duration of the two stories, ' m ' and ' n ' are the total number of shots in the two stories respectively and $\{t_{11}, t_{12}, \dots, t_{1m}\}$ and $\{t_{21}, t_{22}, \dots, t_{2n}\}$ be the duration of shots in stories s_1 and s_2 respectively. Then the criteria for (*visually*) duplicate videos are:

1. $T_{s1} = T_{s2}$
2. $m = n$
3. $\exists i, j$ such that $\psi(c_{1i}, c_{2j}) = 1, \forall 1 \leq i \leq m, 1 \leq j \leq n$

$$4. \quad t_{1i} = t_{2j}$$

The two stories are not duplicates of each other visually, if any of the above condition fails.

The audio patterns or audio fingerprint that are used (generated using audio features such as LPC, MFCC, shifted delta cepstral features etc) are based on perceptual features of audio that are invariant, at least to certain degree, with respect to signal degradations. Thus severely degraded audio still leads to very similar audio fingerprints. These fingerprints are matched for each frame block, which is a group of frames, from two streams. The two streams are duplicates if the fingerprints of all the frame blocks are matched. As the streams can be from different channels, they may not match exactly at the desired points. The match may occur at few samples before or after the desired point. Thus the audio frames are matched in a window of some pre-determined size.

While considerable emphasis has been placed herein on the particular features of this invention, it will be appreciated that various modifications can be made, and that many changes can be made in the preferred embodiments without departing from the principles of the invention. These and other modifications in the nature of the invention or the preferred embodiments will be apparent to those skilled in the art from the disclosure

herein, whereby it is to be distinctly understood that the foregoing descriptive matter is to be interpreted merely as illustrative of the invention and not as a limitation.

Claims:

1. A system for identification, classification, storage, and analysis of news programs, containing an audio channel, video channel, and metadata relating to it, broadcasted/relayed on a television (TV) channel by means of a plurality of TV broadcast streams, said system comprising:

- acquisition module adapted to capture said TV broadcast streams;
- recording module adapted to record said captured streams on a physical storage;
- news program identification module adapted to identify news programs in said stored broadcast streams;
- news program clipping module adapted to separate said identified news programs from other programs;
- advertisement identification module for identification of advertisements from said identified news programs;
- advertisement clipping module adapted for removal of said identified advertisements;
- seam detection module adapted to detect and identify seams of said news programs in order to demark individual stories in a news program;
- keyword generation module adapted to generate a list of keywords;

- text-keyword identification module adapted to identify said created keywords from visual text of identified said news programs;
- speech-keyword identification module adapted to identify the created keywords from the speech of said identified news programs;
- repeat-identification module adapted to identify similar/repeat news programs from said plurality of TV broadcast streams;
- clustering module adapted to cluster said repeat news programs into one news programs, in order to avoid duplication or multiplication;
- removal module adapted to remove said repeat-identified news programs;
- repository adapted to store said news programs and metadata embedded in said news programs; and
- logical interconnection module adapted to logically interconnecting each of said modules for determining the sequence of steps for a multilingual news video analysis system.

2. A system as claimed in claim 1 wherein, said keyword generation module is a multilingual keyword generation module adapted to generate keyword in multiple languages by analyzing the RSS feed making use of the word level dictionaries.

3. A system as claimed in claim 1 wherein, said text-keyword identification module is a multilingual text-keyword identification module adapted to identify said created keywords from visual text of identified said news programs, in different languages, in the visual channel of the news program.

4. A system as claimed in claim 1 wherein, said speech-keyword identification module is a multilingual speech-keyword identification module adapted to identify said created keywords from the speech in different languages, in the audio channel of the news program.

5. A system as claimed in the above claims wherein, said system includes a multilingual lexicon database for generating multilingual synonymous keywords for said created keywords.

6. A system as claimed in claim 1 wherein, said system includes a retrieval module adapted for retrieving a news program from said repository.

7. A system as claimed in claim 1 wherein, said system includes a navigation means adapted for navigation in said repository for retrieving a news program.

8. A repeat-identification module includes visual matching means adapted to use visual cues in order to identify repeat news programs, said visual matching means comprising:

- key frame identification means adapted to identify at least a key frame in a plurality of news program;
- key frame visual feature extruding means adapted to extrude visual features relating to pre-defined parameters of said identified key frames;
- processing means adapted to process said extruded features based on pre-defined tests in order to obtain a processed similarity score in relation to plurality of news program;
- identification means adapted to identify repeat news programs based on pre-defined criteria of said similarity score; and
- deletion means adapted to delete said identified repeat news programs.

9. A repeat-identification module included audio matching means adapted to use audio cues in order to identify repeat news programs, said audio matching means comprising:

- window determination means adapted to determine a window of frames in a plurality of news programs;
- fingerprint detection means adapted to detect audio fingerprint based on pre-defined processing criteria on said determined window of frames;
- processing means adapted to process said detected audio fingerprint based on pre-defined tests in order to obtain a processed similarity score in relation to plurality of news program;
- identification means adapted to identify repeat news programs based on pre-defined criteria of said similarity score; and
- deletion means adapted to delete said identified repeat news programs.

10.A method for identification, classification, storage, and analysis of news programs, containing an audio channel, video channel, and metadata relating to it, broadcasted/relayed on a television (TV) channel by means of a plurality of TV broadcast streams, said method comprising the steps of

- capturing said TV broadcast streams;
- recording said captured streams on a physical storage;
- identifying news programs in said stored broadcast streams;
- separating said identified news programs from other programs;
- identifying advertisements from said identified news programs;

- clipping said identified advertisements;
- detecting and identifying seams of said news programs in order to demark individual stories in a news program;
- generating a list of keywords;
- identifying said created keywords from visual text of identified said news programs;
- identifying the created keywords from the speech of said identified news programs;
- identifying similar/repeat news programs from said plurality of TV broadcast streams;
- clustering said repeat news programs into one news programs, in order to avoid duplication or multiplication;
- removing said repeat-identified news programs;
- storing said news programs and metadata embedded in said news programs; and
- logically interconnecting each of said modules for determining the sequence of steps for a multilingual news video analysis system.

11.A method as claimed in claim 10 wherein, said step of identifying said created keywords from visual text of identified said news programs includes the step of generating keywords in multiple languages.

12.A method as claimed in claim 10 wherein, said step of identifying created text keywords includes the step of identifying created keywords from visual text of identified said news programs, in different languages, in the visual channel of the news program.

13.A method as claimed in claim 10 wherein, said step of identifying speech-keywords includes the step of identifying said created keywords from the speech in different languages, in the audio channel of the news program.

14.A method as claimed in claim 10 wherein, said method includes a step of retrieving a news program from said repository

15.A method as claimed in claim 10 wherein, said method includes a step of navigating in said repository for retrieving a news program.

16.A method as claimed in claim 10 wherein, said step of removing repeat-identified news programs includes a method of using visual cues in order to identify repeat news programs, said method comprising the steps of:

- identifying at least a key frame in a plurality of news program;
- extruding visual features relating to pre-defined parameters of said identified key frames;
- processing said extruded features based on pre-defined tests in order to obtain a processed similarity score in relation to plurality of news program;
- identifying repeat news programs based on pre-defined criteria of said similarity score; and
- deleting said identified repeat news programs.

17.A method as claimed in claim 10 wherein, said step of removing repeat-identified news programs includes a method of using audio cues in order to identify repeat news programs, said method comprising the steps of:

- determining a window of frames in a plurality of news programs;
- detecting audio fingerprint based on pre-defined processing criteria on said determined window of frames;
- processing said detected audio fingerprint based on pre-defined tests in order to obtain a processed similarity score in relation to plurality of news program;

- identifying repeat news programs based on pre-defined criteria of said similarity score; and
- deleting said identified repeat news programs.

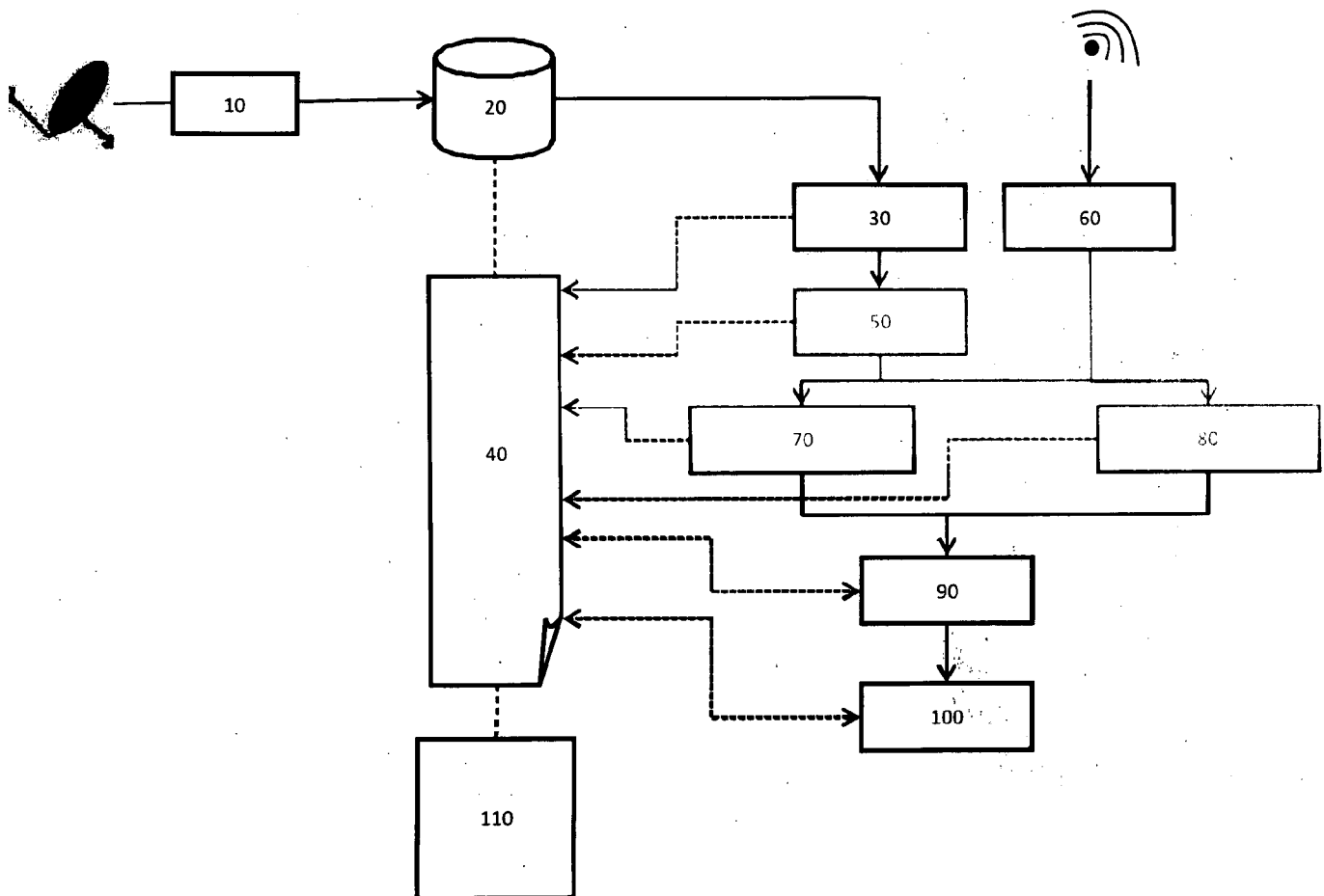


FIGURE 1