Volume 50, Number 4, 2009

ACTA TECHNICA NAPOCENSIS
Electronics and Telecommunications

ASPECTS OF ALGORITHM DEVELOPMENT FOR SYSTEMS WITH ONLINE DETECTION AND RECOGNITION OF TV COMMERCIALS

Radu ARSINTE
Technical University of Cluj-Napoca, Faculty of Electronics, Telecommunications and Information Technology, Communications Department, Tel: +40-264-595699, Str. Baritiu 26-28, Radu.Arsinte@com.utcluj.ro

Abstract: The paper describes a series of research stages in the development process for an automated system, able to detect commercial brakes in TV signals, and recognize a basic set of logos form the TV commercials. First, the conceptual diagram of the system is presented. Theoretical aspects of the recognition process are presented, and there are based on image matching algorithms. The paper presents also the test bench implemented to verify the correctness of the algorithm. The test bench is based on EasyAccess environment from eVision. Using the described test bench it is possible to tune the main parameters of the algorithm and realize extensive tests for different conditions and images. Using this test bench made possible to realize over 600 matching experiments using 20 logos of the most important international brands and 30 images captured in real conditions using different TV tuners.

Key words: *Video acquisition, Commercial brake, Logo Detection, Video Streaming*

I. INTRODUCTION

Detection and identification of TV commercials has an important role in modern embedded systems, like enhanced TVs, or DVD recorders. Applying such features in embedded video systems make possible to add new features, like selective recording or playback, omitting unwanted parts from a desired program or video sequence.

In the literature there are many methods proposed for detection of commercials in TV programs for over 20 years. A general method is to detect a high activity rate (in normal program duration) and detection of the black frames before the commercial break. Such methods have only partial results.

It is possible also to use monochrome transformation of the original image, scene interruption, and action (the number of pixels on the edge which are changing during consecutive frames and the length of the movement vector).

D.W. Blum and a couple of different researchers [1] used black frame detection and activity detector. Activity is the change in luminance level between two different frames.

Commercials are usually rich in activity. When a small amount of movement is detected again, it is considered that the commercial break is over.

Unfortunately, it is difficult to find where is the activity and its duration. Moreover, black frames are also present during normal frames (image dissolution).

Any sequence of black frames, followed by a sequence of significant action could be considered as a commercial break, and eliminated from program.

A different technique, implemented by Iggulden [2] is using the distance between black frames to detect the presence of a commercial break.

Lewine had determined the commercials using the image matching. In a similar manner Forbes uses a video signal identifier to identify similar signals and control the recording.

Nafeh [1] has proposed a method to identify commercials for both audio and video components using neural networks.

John M. Gauch and Abhishek Shivadas are presenting in their works ([2][3]) a method based on the recognition of hidden codes included in the commercials. In this case it is necessary to create previously a database with the features of commercials containing the necessary information. They are presenting an automatic technique to localize unknown commercials, monitoring continuously the TV signals.

The proposed system has two components: detection of repeated video sequence and the classification based on the main characteristics in publicity and non-publicity video sequences.

This system is using temporal video segmentation techniques, adapted for automated partitioning of video sequences based on the semantics of frames and scenes.

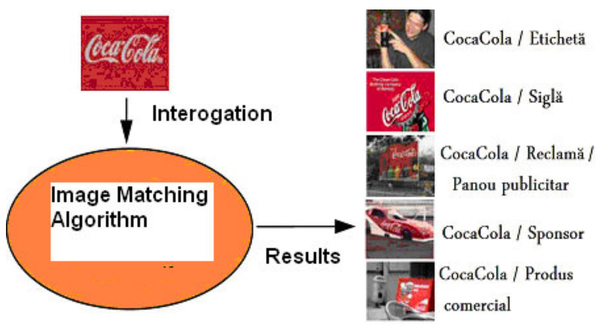

Figure 1. Search based on an image database

Manuscript received October 25, 2009; revised November 18, 2009

During the processing of each video source, the auxiliary information necessary for repeated sequence detection is extracted.

When video transition, marking the end of scene, is detected, all the previous incidences of the clip are detected.

To classify the video sequences in publicity or non-publicity, a number of characteristics are extracted for each video sequence, describing temporal and chromatic variations from the clip.

The authors have three approaches using this information, reaching an accuracy of 93% in new publicity and non-publicity sequences identification during transmission.

Taro Suito, Takao Takahashi, Masashi Ohta, Toshiya Akiba, Noboru Murabayashi, Naohisa Arai şi Masami Tomita invented a commercials detection system, useful for video recording systems like video recorders and DVD recorders.

References ([4]–[8]) are describing different approaches in commercials detection.

II. ALGORITHMS USED FOR DETECTION

We used in this project image matching technique (integrated in eVision), comparing a set of source images, already stored within the system, containing the main characteristics of the commercial video to be analyzed.

Humans are capable to identify easily if an image is similar to another, in a way important to the viewer. This similarity evaluation could be based on shapes, colors, image composition, object proximity or a combination of all mentioned factors.

Computers are less efficient for this task, since are based on a linear logic. Using special methods and systems, it is possible to make the computers able to identify images like human beings.

The basis of any software for image recognition it is the ability to realize adaptive matching, when we are comparing images.

Like in human model of recognition, adaptive matching is giving the possibility to find out if similar, but not identical, images are essentially the same.

Applications realized for identification and image matching are offering the necessary solutions to handle the huge amount of images and video clips necessary to be processed.

The capabilities of systems with interactive technology for image retrieving and recognition are used in systems for image or video sequences archiving.

A system equipped with such features is able to identify commercial logos in images and video sequences.

A realistic approach is to use inter-correlation. To compute the inter-correlation of two windows, a template image is moved pixel by pixel across a larger search window (figure 2), and for each position it is computed the inter-correlation coefficient ρ between template and the corresponding window using the equation (1). The maximum of inter-correlation resulted defines the position of the best matching of the template with the search window.

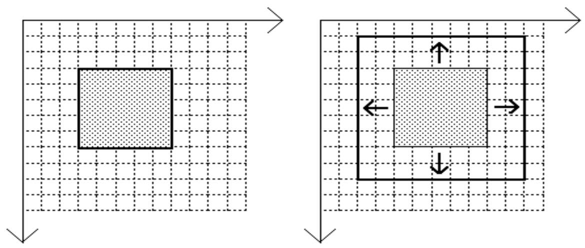

Figure 2. Inter-correlation principle

$$\rho = \frac{\sum_{r=1}^{R}\sum_{c=1}^{C}(g_1(r,c)-\mu_1)(g_2(r,c)-\mu_2)}{\sqrt{\sum_{r=1}^{R}\sum_{c=1}^{C}(g_1(r,c)-\mu_1)^2\sum_{r=1}^{R}\sum_{c=1}^{C}(g_2(r,c)-\mu_2)^2}} \quad ;-1\leq \rho \leq 1 \qquad (1)$$

$g_1(r,c)$ – individual gray levels in template matrix
u_1 – average gray level from template matrix
$g_2(r,c)$ – individual gray levels from searched matrix
u_2 – average gray level from searched matrix
r, c – number of lines and columns from template matrix

III. ALGORITHM TESTBENCH

To test the effectiveness of the matching algorithm we used EasyAcces 6.2 from eVision environment [9]. eVision utility has a image processing library, a large number of functions able to implement different industrial or laboratory applications. It is possible to interface this application with any platform, like Microsoft Basic or Borland Delphi.

The conceptual diagram of the application is presented in figure 3.

The eVision libraries are a set of powerful image processing tools designed for use in industrial vision applications. This set of functions cover most important techniques in digital image processing, from classical algorithms to advanced solutions ready-made for specific tasks.

The key component used in this project is EasyMatch. A typical vision-based application, using eVision, follows a few steps.

1. Pre-processing: the image is possibly modified in order to reduce defects that are unavoidably introduced by the image formation and acquisition stages (geometric distortion, de-focusing, noise) or enhance some desired property such as contrast between objects of interest and background.

2. Location: the objects or parts of interest are roughly or accurately located using various techniques such as segmentation, edge detection or pattern matching. In other cases, the same tools can be used to count the objects if their number is unknown beforehand.

3. Analysis: when relevant locations are determined, processing concentrates on judicious regions of interest and measurements are performed locally: the shape of objects can be quantified by appropriate geometric quantities, defects can be related to abnormal gray level values.

4. Diagnosis: the computed quantities can be used to assess the parts quality and detect defects by comparing them to the expected values for good parts. Alternatively, these quantities can be used to recognize and sort objects, as is the case when unknown markings have to be read.

Once tested, the application can be translated immediately into en embedded application using normal C development tools and eVision libraries.

The eVision libraries are available in two forms, depending on the development platform:

- The C++ interface includes a set of C++ classes and additional global functions. It provides the best efficiency and gives you unlimited access to all eVision features. It can be used from any of the supported C++ compilers.
- The ActiveX interface includes a set of ActiveX controls, i.e. objects equipped with properties and methods. It is very similar to the C++ interface; a one to one correspondence has been established between the C++ classes and ActiveX controls, except for minor details. In certain cases, advanced features of eVision have not been integrated in the ActiveX interface, for simplicity.

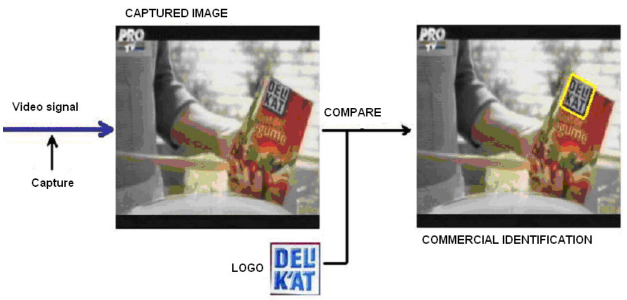

Figure 3. Diagram of logo identification

The most important component used in image recognition is EasyMatch. EasyMatch (figure 4) is a gray level and color pattern matching library. Using this component it is possible to train the system on a reference pattern and afterwards locate its occurrences in other images. This tool is very convenient when the position of a given part in the field of view is unknown or if the presence of parts must be controlled. The library works by using correlation, i.e. measuring discrepancies between the pattern and the target image.

The following features are provided in this module:

- multiple pattern occurrences: several occurrences of a pattern, up to a user-defined number, are returned. Only the reliable ones are retained.
- standard, offset-normalized, gain-normalized and fully normalized correlation: the correlation is computed on continuous tone values (as opposed to binary). It is well known that when the lighting conditions vary, as it is often the case, straight comparison of the pattern and image behaves badly. To cope with this, automatically adjusting the contrast and/or intensity of the pattern before comparison is very effective. This process is known as normalization. EasyAccess provides four distinct normalization modes, depending on whether a gain and/or offset compensation is used.
- translation, rotation and isotropic/anisotropic scaling: to find the best matches between the pattern and target image, the target is allowed to translate horizontally and vertically. Additionally, it can be allowed to rotate and/or to change its scale in the X and Y directions simultaneously or independently. The rotation angle and scale factors vary in a user-specified interval. All degrees of freedom can be combined at will.
- variable accuracy, up to sub-pixel level: the accuracy with which the pattern is measured can be chosen (the less accurate, the faster). By default, the position parameters for each degree of freedom are computed with a precision corresponding to the size of a pixel. Lower precision can be enforced. On the other hand, one tenth-of-a-pixel accuracy can be achieved.

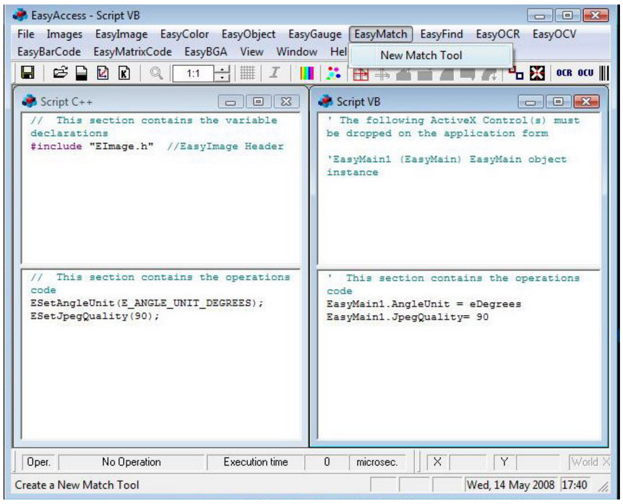

Figure 4. EasyMatch main screen

- set of don't care pixels: when the pattern cannot be inscribed in a rectangular region of interest, the surrounding of the pattern can be ignored by setting the pixels values below a threshold level. These pixels will not take part in the matching process. The same feature can be used if parts of the template change from sample to sample.
- handling gray level and color images: EasyMatch processes BW8 images as well as C24.
- works with non-square pixels: when images are acquired with non-square pixels, rotated objects appear skewed. Taking the pixel aspect ratio into account can compensate for this effect.

IV. EXPERIMENTAL RESULTS

To verify the algorithm performance we used 30 images, captured from real TV sources and 20 images containing logos for products and services. We have realized more than 600 matching experiments. The results are presented in the following tables (in a condensed form). In table 1 the winning logo (Coca-Cola) is very close to the next competitor (Vodafone) (figure 5), so the decision should be based on additional criteria to have a correct identification. The dominant color (hot red - used as a matching criterion in both cases) seems to be the explanation in this case.

A different conclusion is that medium scores (0,6-0,7) are not significant in logo recognition, and should be avoided as a unique recognition criterion. A value over 0.75 is marking with a high degree of confidence a real match. Each table contains the name of the reference logo, correlation score, coordinates for the center of the identified part in image, and the scaling factor. Two sample tables generated for two test images, in the extensive testing process are presented in Tables 1 and 2.

Table 1.

Logo name	Score	Center X	Center Y	Scale %
Coca-Cola	**0.6200**	**126.00**	**575.00**	**81.36**
DeliKat	0.4383	794.00	430.00	7.56
Blend-a-Med	0.1819	817.00	707.00	36.78
Germanos	0.3797	762.00	424.00	42.70
Oust	0.2234	720.00	301.00	62.22
Pronto	0.4335	209.00	447.00	70.50
Fiat	0.3612	294.00	435.00	54.73
Schweppes	0.2831	685.00	196.00	58.44
Renault	0.5586	617.00	576.00	65.30
Vodafone	0.5900	660.00	546.00	52.34
Nivea	0.3452	753.00	445.00	66.87
7 Days	0.5758	704.00	377.00	11.64
Nokia	0.0854	815.00	340.00	67.49
Milli	0.4821	771.00	429.00	78.80
Ford	0.1695	831.00	428.00	65.99
Cappy	0.1694	740.00	207.00	49.88
Lays	0.3855	511.00	153.00	70.12
BCR	0.0883	790.00	418.00	76.85
Borsec	0.1729	793.00	422.00	81.75
ROM	0.5823	862.00	265.00	59.54

Table 2

Logo name	Score	Center X	Center Y	Scale %
Coca-Cola	0.2737	543.00	424.00	64.84
DeliKat	0.6264	419.00	610.00	4.39
Blend-a-Med	0.3480	934.00	704.00	57.82
Germanos	0.2304	675.00	520.00	64.47
Oust	0.2769	564.00	542.00	36.20
Pronto	0.5534	71.00	194.00	59.83
Fiat	**0.8850**	**525.00**	**445.00**	**95.81**
Schweppes	0.3096	530.00	428.00	56.68
Renault	0.2679	112.00	70.00	35.34
Vodafone	0.2576	959.00	650.00	58.63
Nivea	0.3724	970.00	692.00	55.37
7 Days	0.2563	518.00	432.00	73.82
Nokia	0.3718	406.00	123.00	88.30
Milli	0.3679	750.00	210.00	54.52
Ford	0.1627	546.00	396.00	98.94
Cappy	0.3916	401.00	185.00	49.74
Lays	0.1779	360.00	346.00	72.89
BCR	0.1483	286.00	268.00	49.91
Borsec	0.3381	672.00	88.00	97.11
ROM	0.4812	109.00	78.00	52.52

Table 2 shows a winner (Fiat) clearly marked compared with the next competitor.

We have repeated the test for all the 30 images captured from normal TV stations using standard TV analog tuners and digital devices like Satellite Boards. For digital sources, it is possible to avoid the image pre-processing stages, since the native quality of images is improved. The full set of results is presented in [10].

From the presented results, it is possible to observe that we have good correlation scores for 16 images. The mentioned values are around 0.7 and above. This fact is indicating that, from a series of 30 captured images, with 20 images with publicity material we identified correctly 16 logos.

This fact is confirmed by the frame generated after

Figure 5. Two logos with similar scores (Table 1)

recognition, indicating accurately the identified zone where the logo is located in the target image (figure 6). For 4 of 20 images with commercial content, the scores are not satisfactory, without correctly identified images (still present in the image). A first conclusion is that the precision of the program is about 80%. The rest of 20% percents with undetected logos, are for captured images with low inherent quality (noise and reflections in the target image).

On the other hand the reference image, containing the product logo, must have the best possible quality, and the logo must be clearly marked in the reference image.

The precision of the program can be improved using high quality images for the target image. This fact is observed easily for images captured from digital transmissions (DVB).

For a better detection scaling factor of the image was modified from 50% to 300%. The normalized correlation is used in most cases, but for some images it was necessary to change the type of correlation. In most test images the logos are presented in the right position, but it is possible to set the program parameters to modify the angle, and detect the tilted logo.

A disadvantage of EasyMatch module is the fact that it is possible to work in a session with only two images (reference and target). This characteristic results in a significantly increase of executed processes. It is necessary to have a matching experiment for each logo, in

Figure 6. Marking the recognized image in source image

images with multiple logos. In our case it was necessary to have around 600 experiments. Execution time is another important issue, to evaluate if the target system will be able to identify promptly the logos. In our series of experiments processing time is variable in 20-2700 ms interval. Execution time is increasing with the necessity of scaling of the reference image (logo). The experiments were realized on computers using normal Pentium IV architectures (desktop and laptop versions).

V. CONCLUSIONS

This series of experiments has multiple theoretical and practical conclusions.

The complexity of the eVision program is an excellent opportunity to verify extensively the algorithms before the implementation. But this complexity is a major handicap in real implementations, where processing time is the most important issue for real time data processing. It is easy to see from the results of the experiments that using the same processing power it will be possible to recognize a logo image in 3-5 seconds, so frame by frame analysis seems to be impossible. A practical solution is to use different implementations of the recognition algorithm (eventually optimized in assembler) to have a near real time analysis system.

A different practical solution could be to detect the commercial break, using one of the methods already presented in the first paragraph, to acquire and store the information from the commercial brake, and detect later the logos, during normal TV program.

For a specific logo the program can be easily modified or adapted, in this case it is possible to combine the matching parameters with additional features, for a better detection.

The next stages of the experiment will be necessary to establish the best algorithms and parameters useful for a specific detection scenario. A special attention will be given for preprocessing algorithms, necessary when the input image doesn't have a good quality. This aspect will be explored in a future paper.

Practical implementation will be done first in a normal PC architecture, but the final goal is to try a real embedded solution using multimedia DSPs.

ACKNOWLEDGEMENT

This work has been supported by the grant project CNCSIS No. 339/2007.

REFERENCES

[1] N. Dimitrova, S. Jeannin, J. Nesvadba, T. McGee, L. Agnihotri, G. Mekenkamp – "Real-Time Commercial Detection Using MPEG Features", Invited paper, *Proc. 9th Int. Conf. On Information Processing and Management of Uncertainty in knowledge-based systems (IPMU 2002),* pp. 481-486, Annecy, France

[2] Abhishek Shivadas, John M. Gauch – "Real-Time Commercial Recognition Using Color Moments and Hashing", *Proceedings of International Workshop on Video Processing and Recognition, Fourth Canadian Conference on Computer and Robot Vision (CRV 2007),* pp.465-472

[3] John M. Gauch, Abhishek Shivadas - *Finding and identifying unknown commercials using repeated video sequence detection*, Elsevier Science Inc. New York, NY, USA, July 2006

[4] Michael Joseph DeLuca - *Selective commercial detector and eliminator apparatus and method*, US Patent, Issued on October 26, 1999

[5] Xian-Sheng Hua, Lie Lu, Mingjing Li, Hong-Jiang Zhang - *Learning-based automatic commercial content detection*, US Patent, Issued on January 16, 2007

[6] Eric Armengaud - *Commercial-break detection device*, US Patent Issued on February 27, 2007

[7] Masami Mizutani, Shahram Ebadollahi, Shih-Fu Chang – "Commercial detection in heterogeneous video streams using fused multi-modal and temporal features", *Proceedings of(IEEE International Conference Acoustics, Speech, and Signal Processing* (*ICASSP'05)*, pp.157-160

[8] Alberto Albiol, María José Ch. Fullà, Antonio Albiol, Luis Torres – "Commercial detection using HMMs, Image Analysis for Multimedia Interactive Services", *Proceedings of Wiamis'2004, Lisboa*, Portugal, 21-23 april 2004.

[9] * * * - *eVision - User's Guide*, Euresys , 2003

[10] Angela Sarghie – *Analiza on line a semnalului video cu aplicații în monitorizarea duratei de reclama a canalelor TV,* Diploma Thesis, Scientific superviser Prof.Dr.Ing.Radu Arsinte, Technical University of Cluj-Napoca, Romania, 2008