Available online at www.sciencedirect.com
ScienceDirect
Computer Standards & Interfaces 29 (2007) 335 – 342
COMPUTER STANDARDS & INTERFACES
www.elsevier.com/locate/csi

A real-time shot cut detector: Hardware implementation

Lotfi Boussaid [a,c,*], Abdellatif Mtibaa [b], Mohamed Abid [a], Michel Paindavoine [c]

[a] Ecole Nationale d'Ingénieurs de Sfax: Unité de recherche CES, France
[b] Ecole Nationale d'Ingénieurs de Monastir: LaboratoireEμE, France
[c] Laboratoire LE2i, France

Accepted 24 May 2006
Available online 24 July 2006

Abstract

With the enormous growth in digital audiovisual (**AV**) information in our life, there is an important need for tools which enable describing the **AV** content information. In this context, the **MPEG-7** standard was developed in order to provide a set of standardized description tools which generate metadata about **AV** content. However, before any content-based manipulations, the hierarchical structure of video must be determined. This process is known as shot boundary detection or in other case scene change detection. In this paper, an old and reliable method based on local histogram has been used to implement shot cut detector for real-time applications. Since software implementation on **PC** is not suitable for this algorithm due to the sequential treatments of the processor, we have used an **FPGA**-based platform.

Keywords: Shot boundary detection; Video segmentation; Local histogram; **MPEG-7**; **FPGA**; Virtex XCV800; Hardware implementation

1. Introduction

Video data is becoming very important in many application domains such as digital broadcasting, interactive-**TV**, video-on-demand, computer-based training, and multi-media processing tools. Furthermore the development of the hardware technology and communications infrastructure has made automatic analyzing of video content very challenging.

In this work, which is part of a project thesis, we present the different steps of the hardware implementation of shot cut detector based on local histogram algorithm. This old and reliable approach is a descriptor in the **MPEG-7** standard. The objective of the **MPEG-7** ("Multimedia Content Description Interface") standard is to specify a standard set of descriptors and description schemes for describing the content of **AV** information. It specially standardizes a number of description tools which describe **AV** content ranging from low-level features to high level semantic information. In other words, it provides a set of standardized description tools which generate metadata (data about data) about **AV** content by extracting information of interest from it, to facilitate a variety of applications including image and scene retrieval [1]. In this context, the local histogram approach constitutes a low-level feature which is utilized for video segmentation and image and scene retrieval. In order to develop any content-based manipulations on video information, hierarchical structure must be determined. In this way, a standard hierarchical video model was defined as shown in Fig. 1. This model is composed of some elementary units as scenes, shots, and frames. In this structure a shot is defined as an unbroken sequence of frames from a single camera, where a scene is a set of shots with semantic link, location unit and action unit [2].

In produced video such as television or movies, shots are separated by different types of transitions, or boundaries. Although well known video editing programs such as Adobe Premiere or Ulead Media Studio provide more than 100 different types of edits, we classify in general transition effects into two categories [3]. The simplest transition is a cut, an abrupt shot change that occurs between two consecutive frames. Gradual transition such as fades and dissolves are more complex. Shot boundaries are fades when the frames of the shot gradually change

* Corresponding author. Ecole Nationale d'Ingénieurs de Sfax: Unité de recherche CES, France.
E-mail addresses: lotfiboussaid@yahoo.fr (L. Boussaid), Abdellatif.mtibaa@enim.rnu.tn (A. Mtibaa), mohamed.abid@enis.rnu.tn (M. Abid), paindav@u-bourgone.fr (M. Paindavoine).

0920-5489/$ - see front matter
doi:10.1016/j.csi.2006.05.009

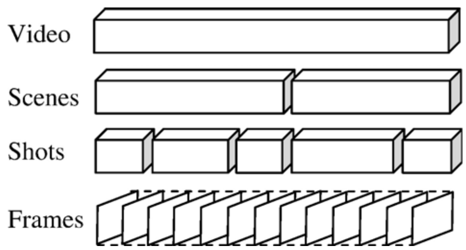

Fig. 1. Standard hierarchical video model.

from or to black, and can be dissolved when the frames of the first shot are gradually morphed into the frames of the second [4]. Fig. 2 shows an example of transition effects.

Most of the existing methods of video segmentation have to challenge the difficulty of finding shot boundaries in the presence of a camera or object motion and illumination variations which can lead to false detection. In other cases, frames that have different structures but similar color distributions can give a missed detection [5]. The study of the state of the art shows that several methods for **SBD** were proposed. These methods can operate in different environments such as temporal, frequency, uncompressed and compressed domains.

On the other hand, Dailianas and Lefèvre [6,7] have distinguished two classes of methods: Those which could be done off-line and have high complexity, and others which are dedicated for real-time applications. In this case, some constraints have to be taken into account. In this paper we have used an old and reliable method based on local histogram and proposed by Nagasaka and Tanaka [8]. They have divided each frame into 16 blocks and computed local histogram before evaluating a difference metric. Histogram-based methods have shown a good performance for shot cut detection.

To operate in real-time condition, computational time of the difference metric mustn't exceed the blanking time which is about 2 ms. Since software implementation on **PC** is not suitable for the local histogram algorithm due to the serial architecture of the microprocessor, we have designed our system on a hardware platform based on Virtex xcv800 **FPGA**.

This paper is organized as follows. In Section 2 we present the different methods proposed for the detection of the abrupt shot changes. Section 3 describes the specifications of the local histogram method which we tested on a set of video sequences, in different color spaces, different types of quantization and different formats of sub-sampling. The concept of the hardware design and the interpretation of the hardware implementation results are presented in Section 4. Finally Section 5 brings the conclusions and the future works.

2. Related work

An important variety of shot boundary detection algorithms was proposed in the last decade. The study of the current state of the art shows that we can classify these algorithms into three generations. The first generation concerns methods which measure distance of similarity between adjacent frames by using elementary features extraction such as pixel differences, global and local histogram differences, motion compensated pixel differences and **DCT** coefficient differences [9–12]. In the second generation, hybrids of the above methods have been investigated [13–15]. In this way J. S. Boreczky [16] has combined audio and video features in a hidden Markov model to increase the shot change detection. Although these techniques have brought improvement to the quality of shot change detection, they increased complexity and computational time. The most recent algorithms have introduced intelligent algorithms such as fuzzy approaches and those based on neural network [17,18].

To compare the ability of real-time or close to real-time implementation of the different methods, Dailianas [6] has evaluated the complexity of many approaches by estimating the number of operations when measuring dissimilarity between two consecutive frames. In this way, he has used an assumption that addition, subtraction and multiplication require time equivalent to one operation, whereas divisions take approximately four times

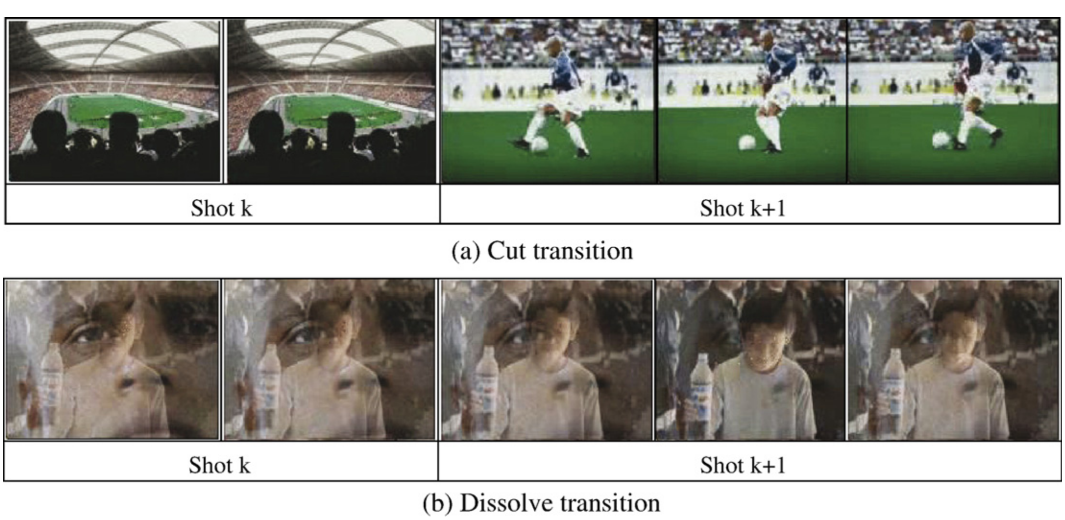

Fig. 2. Example of transition effects.