UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR

PROYECTO FIN DE CARRERA

CLASIFICACIÓN AUTOMÁTICA DE VÍDEOS POR GÉNERO

Loreto Felipe Sánchez-Infante

SEPTIEMBRE 2011

CLASIFICACIÓN AUTOMÁTICA DE VÍDEOS POR GÉNERO

AUTOR: Loreto Felipe Sánchez-Infante
TUTOR: José María Martínez Sánchez

Email: {Loreto.Felipe, Josem.Martinez} @uam.es

VPULab (Grupo de Tratamiento e Interpretación de Vídeo)
Dpto. de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Septiembre de 2011

RESUMEN

Debido al desarrollo de mejoras en la digitalización de contenido multimedia, hoy en día se producen y distribuyen cantidades tan grandes de datos que comienza a ser indispensable la utilización de herramientas de clasificación automática que, eficientemente, ayuden al usuario a encontrar únicamente los contenidos que son de su interés.

Este tipo de progresos aceleran la detección de los contenidos deseados, consiguiendo además un ahorro considerable de recursos y tiempo. Puede ser también de gran utilidad para detectar vídeos de contenido violento o pornográfico en canales como Youtube donde deben ser eliminados.

El trabajo estudiado durante el estado del arte se ha centrado principalmente en la extracción de descriptores de tres tipos: texto, audio y visual. El sistema que se ha desarrollado en este trabajo obtiene únicamente características visuales para diferenciar entre los diferentes vídeos. Estos parámetros son posteriormente utilizados para clasificar 8 clases de deportes: fútbol, baloncesto, tenis, vóley playa, snowboard, deportes acuáticos, fórmula 1 y ciclismo. Investigaciones posteriores en esta área deben centrarse en aumentar el rango de géneros de vídeo que pueden diferenciarse, por ejemplo, al ámbito de los géneros televisivos (como dibujos animados, noticias, anuncios…) y en la mejora de las técnicas utilizadas hasta ahora a través de reducir los requerimientos computacionales.

Palabras clave: Clasificación automática vídeo * Reconocimiento de género * Extracción descriptores * Deportes

ABSTRACT

Due to the improvements on the digitalization of multimedia data, the volume of videos that we can produce and distribute has become so huge that the need of effective tools to automatically classify them to help the viewers to find videos of their interest is indispensable. In this context, automatic video classification by genre can provide a simple and efficient solution to organize these multimedia contents in a structured and intuitive way.

This kind of progress accelerates the retrieval of the desired contents, achieving a considerable save of resources and time. It can also be useful to detect videos with violent or pornographic content which should be discarded immediately in channels like Youtube.

Previous methods have focused on three types of feature extraction: text, audio and visual classification. The algorithm developed in this work makes use only of visual features to classificate the different genres. These features are used to classify videos among 8 common sports: football, basketball, tennis, beach volley, snowboard, water sports, formula one and cycling. Further research should focus on widening the range of possible genres to be distinguished for example, TV genres such as cartoons, news, commercials…, and on the improvement of the existing techniques by using techniques with lower computational requirements.

Keywords: Automatic video classification * Genre recognition * Feature extraction * Sports

AGRADECIMIENTOS

Para comenzar, me gustaría dar las gracias a mi tutor José María Martínez por brindarme la oportunidad de realizar este proyecto en el VPU-Lab. Gracias por el apoyo y la paciencia, por los incontables emails, por los riquísimos tomates…sin tu orientación no hubiera sido capaz de terminar este proyecto.

También me gustaría destacar la ayuda tan valiosa de Víctor Valdés, que desde el principio no sólo afianzó los cimientos de este trabajo sino que siempre ha estado ahí para ofrecerme su ayuda y asesoramiento, así como otros miembros del laboratorio, Javier Molina y Juan Carlos San Miguel, que siempre han tenido un momento para atender mis dudas al instante por muy liados que estuviesen. A todos vosotros, gracias.

Acercándose el final de una etapa de mi vida tan importante, me gustaría también dar las gracias a todos los profesores que durante tantos años nos han visto crecer, no sólo académicamente sino también como personas. Ha sido todo un privilegio haber podido aprender tanto de vosotros. En especial, gracias a Jesús Bescós (porque nunca pensé que el suspenso de EyC hubiera podido ser el preludio de tantos consejos y motivaciones), Doroteo Torre (porque es un placer haber podido asistir a tus clases), Jorge López de Vergara (por sus consejos como jefe de la ORI), Eloy Anguiano (por las innumerables charlas) y, una vez más, a José María Martínez (porque también como profesor tuvo que sufrir mis emails..).

A mis compañeros (muchos de ellos amigos) de universidad, que les ha tocado avanzar, adelantarme, quedarse atrás, estudiar, prácticas, no estudiar, ir a clase, más prácticas, y demás hazañas conmigo, muchas gracias. Estos cinco años se me han pasado volando gracias a todos los buenos momentos que hemos compartido juntos. Por las mañanas sufridas alternando biblioteca y laboratorios, por las tardes de mus que desgraciadamente se fueron reduciendo a medida que avanzábamos, por las noches de Carcassone y Munchkin; porque Piticli no es sólo un pájaro, sino dos; por enseñarme que Alcobendas y Algete no son pueblos, sino ciudades; por deleitarme con maravillosas croquetas caseras y descubrirme los secretos del baloncesto; por los sudokus; por tantas y tantas noches juntas; por darme un papel principal en “La guerra de las galaxias”; por hacerme sobrepasar la barrera de los 200; por dejarme la mejor cama si hubiera ido a Barcelona; por los soles; por todas las cervezas, las fiestas por Madrid y los tarancones… y por muchas cosas más, gracias a todos, os deseo lo mejor.

A Luis, porque cada meta que alcanzo es un abrazo más que me debes. Porque fuiste, eres y serás la razón de seguir adelante, superándome. “Porque el sol siempre brilla, aunque no se vea”. Gracias por tus ánimos y tus tardes de pardos. Siempre te llevo conmigo.

A Irene, por ser amiga y después jefa. Gracias por tu apoyo y tus consejos, antes y después de trabajar en la ORI. A Rubén y Lolo, porque guardo con mucho cariño todos los momentos juntos. No sólo sois grandes deportistas, sino aún mejores personas. Ha sido un verdadero placer trabajar con vosotros.

A mis amigos del Erasmus, por hacer que jamás pueda olvidar Berlín. En especial a Coco, por ser la mejor “Tochter” posible, y a Lorena y Samu, con los que espero seguir compartiendo barrio muchos años más. Vielen Dank!! :)

A mi familia, porque no sería nada sin vosotros. Gracias por haberme mimado, apoyado, arropado y atendido, incluso cuando decidí la locura de meterme a Teleco. A mi madre, Marta, y a mi padre, Luis, por servirme de ejemplo durante toda mi vida, porque no existe nada que no pueda agradeceros, por estar siempre ahí, en los buenos y en los malos momentos. Gracias a ambos por hacerme sentir la hija más afortunada del mundo. A mi hermana Elena, porque las riñas y las risas contigo siempre me han servido para hacerme más fuerte. No habría llegado hasta aquí sin tu ayuda. A mis tíos, primos y abuelos, gracias por vuestra confianza y dedicación.

Gracias a la familia Jiménez Martínez por arroparme como si yo también compartiera estos apellidos.

Porque ya no recuerdo lo que es no tenerte a mi lado. Porque siempre consigues sacarme una sonrisa cuando más lo necesito. Porque eres mi apoyo, mi sustento, mi confidente, mi mejor amigo, mi sol, mi todo. Porque siempre has creído en mí, incluso cuando yo no confiaba. Porque siempre me has cuidado y me gustaría que lo siguieras haciendo siempre. Por acompañarme hasta el infinito y más allá. Gracias Javi. Esto también es tuyo.

Son muchas las personas especiales a las que me gustaría agradecerles su amistad y su ánimo durante cada una de las etapas vividas y por vivir. Algunas están aquí conmigo y otras en mis recuerdos y en mi corazón. Sin importar dónde estén, o si alguna vez llegarán a leer estas líneas, quiero dedicarles este trabajo como agradecimiento por haberme acompañado a cruzar con firmeza el camino de la superación.

A TODOS VOSOTROS, MUCHAS GRACIAS.

INDEX

TABLE INDEX

FIGURE INDEX

CHAPTER 1. INTRODUCTION

1.1. INTRODUCTION TO AUTOMATIC VIDEO CLASSIFIERS

The technological advances of this era have led society to make the digitization of information a necessity. The low costs of electronic devices and the fact that most of them offer the possibility of Internet access and data storage, have catapulted the tendency to transform documents, music, newspapers, etc. from paper to media to exponential growth. This modernization can be summarized in a significant increase in the amount of data in digital format that any user can access and, therefore, to a greater difficulty to access just the desired information.

The cataloging of all this content is a tedious task which requires not only a great human effort, but also a considerable expenditure of time. This is why it becomes imperative to have a method to automatically classify and sort all this content. Thus, it is faster and easier for users to find only the data of their interest according to a specific category (e.g. genre, places, people).

The variety of multimedia materials is so broad that the classification by genre is considered to be an effective way to manage such content. For databases that store movies or videos of any kind, the use of these techniques can be particularly useful. The main challenge is to extract a reduced number of features which are representative enough to distinguish between genres.

This work is focused on the automatic classification of the genre "sports video summaries". In particular, to validate the proposed approach, an experimental framework capable of discerning between videos summaries of *football, basketball, tennis, beach volley, snowboarding, water sports, Formula 1* and *cycling* has been developed.

1.2. MOTIVATION AND OBJECTIVES

Video automatic classifiers avoid human intervention, what can be really useful to speed up the processing of huge amounts of data.

The aim of this work is to develop a framework that is capable of classifying automatically a number of videos specifying the genre to which they belong. The experience of the VPULab on competitions such as TRECVID and the advances on video summarization has been the motivation to proceed with this project. The basis of this work relies on the work on on-line video abstract generation of multimedia news [1] from which most of the feature extraction methods have been reused.

The objective of this Master Thesis project is to adapt an algorithm developed to generate multimedia news from on-line videos to another framework capable of classifying 8 different types of sport video summaries. The study of the viability of the system, analyzing its characteristics according to the technology available at this moment, has been always the main issue during the whole analysis. It is important to consider that every single video used to this purpose has been previously summarized on a pre-processing phase. This way, we obtain

what we call “Key frames”, which are representative images selected every 25 seconds, from which our descriptors will be extracted.

To sum up, to carry out this project our methodology has focused on:

- Build up a robust database so that both, training and testing, would give us relevant information. In our case, video summaries have been extracted from Youtube.

- Gain enough experience to develop a possible solution to the automatic genre classification problem. We have based our work on the previous systems.

- Develop an evaluating phase enough detailed to obtain conclusions about the viability, the drawbacks and the efficiency of our system.

Improvements on this area should extend the application of these techniques not only to genres completely different but also to distinguish between action movies and comedies, for example.

1.3. DOCUMENT DESCRIPTION

This document is structured as follows: after this introduction, Chapter 2 presents the state of the art related to the classification of sports videos. This includes not only an overview of the different techniques studied up to the present time, but also a detailed explanation on why have we chosen SVM for our system. Chapter 3 describes the decisions that have been made to optimize the system due to a specific flow of design stages. This section includes a description of each of the selected features that are going to be used. Chapter 4 presents the evaluation part, this means, taking conclusions about the experiments that have been done in each of the simulation phases. The detailed results are included in the Appendixes. Finally, chapter 5 details future work in foreseen investigation and the conclusions about the final results and the effectiveness of the system.

CHAPTER 2. STATE OF THE ART

2.1. INTRODUCTION

The automatic classification of video is a recent initiative originated by the massive availability and use of digital content. Currently, the general interest of researchers for the development of new techniques in this area is growing progressively. Many investigations have attempted to find the optimal classifier during decades, but the reality is that video classification is not an easy task. Some of these techniques can be found in the following section.

The main problem of genre classification assumes that it is a limited concept, that is, its definition applied to multimedia content ranges from social, cultural or historical to sports, movies, cartoons or even violent and pornographic content. This drawback is compounded by the problem that these ratings are usually done on a subjective level what means that the same video can be classified by two different people into two different genres according to their opinion. So far, one of the advantages of automatic video classification is that the distinction between genres is ruled by the characteristics obtained during the training, in fact, by numbers, which give to this technique and, upon some extend, an objective point of view.

In the sports context, on which this project will be based, we can define *genre* as those features that can distinguish one video of football from another of basketball, for example. Our job is to "teach" the "machine" to be able to differentiate the bounds that categorize each class individually. Genre classification involves two steps: feature extraction and data classification. Feature selection is a crucial step to reduce the complexity of the system and must be taken into consideration before the start of the classification. This set of features should be discriminant enough to emphasize the characteristics of each genre while preserving relatively reduced in number so that too many values will not lead to confusion between genres.

Genre labeling is the easiest way to retrieve a particular video among the immensity of a collection. Still, it is important to note that it is not a task that is performed simply, but is usually heavy and unspecific. Therefore, a specific taxonomy of genres should be found to ensure an efficient classification which can be easily understandable by the user, and of course, which requires an extraction of features sufficiently representative.

This section reviews different techniques from previous research on automatic video classification and discusses the particularities of some of the most common genres in video classification, including sports genre which concerns us. The last section of the chapter makes an introduction to the machine learning technique chosen for the classification: the *Support Vector Machine*.

2.2. VIDEO GENRE CLASSIFICATION MODELS

The development of a new classifier requires necessarily the study of previous systems to learn which characteristics will acquire the best results for our specific purpose. This section makes an overview of some of the different techniques studied during the implementation of this work.

The first references of automatic video classification date from 1995 by Fischer et al. [2]. The developed system attempted to classify five different genres (tennis, car-racing, news, commercials and cartoons) and consisted of 3 stages: the first one, obtained the syntactic properties of the video (primarily visual and audio statistics), the second one characterized the style attributes of those properties and, finally, a specific profile was determined according to these properties to be compared with well-known profiles, each one typical from a particular genre. Several further approaches to solve the genre classification problem led the classification techniques to the discrimination of one single genre at the time.

In the literature we can find many different types of classifications (see section 2.3.) depending on the classifier used in each of them. For example, Dimitrova et al. [3] employed a hidden Markov model (HMM) classifier to distinguish between four TV genres using face and text tracking. This classifier was also used to determine the structural characteristics of soap operas, comedies and sport videos by Taskiran et al. [4] and to distinguish between cartoons, commercials, weather forecasts, news and sports by Liu et al [5].

Another model used in video classification is the Gaussian mixture model (GMM) classifier. It is common to be used to extract motion information. Roach et al. [6, 7] classified five genres (sports, cartoons, news commercials and music) using only video dynamics. They adopted an approach based on background camera motion, foreground object motion and aural features. This system was simple but efficient: it achieved detection errors below 6%. Xu et al. [8] considered the same five genres and the same classifier but used a principal component analysis (PCA) on spatio-temporal audiovisual feature vectors.

These features were also extracted in the classification system by Yuan et al. [9] using a hierarchical support vector machine (SVM). Temporal features concern parameters such as shot length, cut percentage, average color difference and camera motion. At spatial level can be found features such as face frames ration, average brightness and color entropy. The set of features was carefully selected to suit the classification of movies and sports. The hierarchical SVM provides a subgenre division: Movies can be classified as action, comedy, horror or cartoon and sports as baseball, football, volleyball, tennis, basketball and soccer. Sports categorization achieved 97% of precision.

Parallel neural networks (PNN) are also very useful in classification terms. [10] combines several types of content descriptors to distinguish between seven video genres: football, cartoons, music, weather forecast, newscast, talk shows and commercials. The extracted features can be grouped in four information blocks: Visual-perceptual (color, texture and motion), structural (shot, length, shot distribution, clusters and rhythm and saturation), cognitive (face properties) and aural (text and sound characteristics) descriptors. The accuracy rates up to 95%. The same genres were classified in [11] extracting three categories of content descriptors: temporal (action content and transitions), color (color distribution and properties) and structural (contour segments) level. The novelty of this work relies on the computation of color parameters and contour information which have not yet been exploited with existing genre classification approaches.

In this work we propose six categories of content descriptors (Face Detection, Color Variety, Frame differences, Shot variation, DCT coefficients Energy and Image Intensity) to classify eight different categories of sports: football, basketball, tennis, beach volley, snowboard, water sports, formula one and cycling. The framework explained in this work has adapted the extraction of some of the features from [1] so that the integration in the summarization would have the minimum cost.

2.3. CLASSIFICATION TECHNIQUES

Talking about techniques related to the classification of large amounts of data, there are two concepts which are always remarkable: *Data Mining* and *Machine Learning*. The first one obtains patterns or models from the data collected. The second one, Machine Learning, is based on the comparison of the similarities that any type of classifier may have.

We can distinguish different types of classifications depending on the characteristics and criteria that they have as it is shown on the following figure [12, 13]:

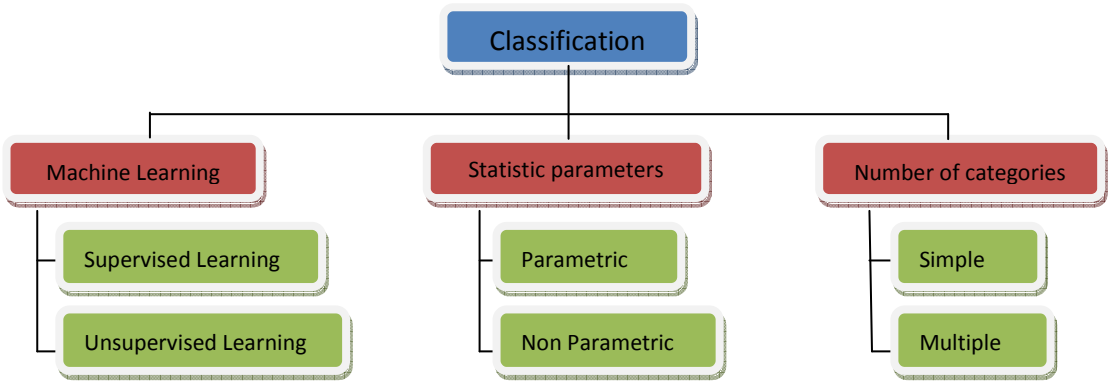

Figure 1- Types of classification

According to the figure above, next we will explain the different possibilities for our classification:

a) Machine Learning:

- ***Supervised learning:*** Some examples are manually labeled in advance so that a similarity relationship can be established between the inputs and desired outputs of the system. This training is not always possible, because it needs to have an expected output.

- ***Unsupervised learning:*** In this case there is no previous database for the training phase, so there are no patterns to learn from, this is, the machine learns automatically as the inputs come. The main drawback is making final decisions among all the different patterns. Inputs will be considered as a set of random variables, building a model of density for the data set. It is commonly known as *Clustering*.

b) Statistic Parameters:

- ***Parametric:*** In order to estimate and learn the unknown parameters a well known statistical model is used, such as the probability density function.

- ***Non Parametric:*** It is divided into two types: based on patterns or based on examples. The first one obtains a description of each category using a vector of features. The classification of an object is made according to the similarities between its vector and the pattern of each category. An example would be the Rocchio classifier [14]. The second type is based on the similarities with some examples taken from the set of training. KNN, K- Nearest Neighbor would belong to this kind.

c) Number of categories:

- ***Simple:*** Only one category is classified. It detects which objects belong to this single category. It can be seen as a binary classification that takes 1 if corresponds to the chosen category or 0 if it is more similar to one of the others.

- ***Multiple:*** Any category can be classified. The same object may have similarities to more than one category. In this classification is selected the most similar one.

Once we know the different types of characteristics that a classification may have, in the following section we will explain some of the most important techniques used in this context.

2.3.1. K-Nearest Neighbor (K-NN)

This is a nonparametric supervised classification method based on a set of prototypes and examples previously used to estimate the value of the density function F(x|Ci), that is, calculate the probability that the value x belongs to class Cj .The values are multidimensional vectors such as $x_i = (x_{1i}, x_{2i}, ..., x_{pi}) \in X$. To find the nearest neighbor, the distance between the different training examples and the x value in the space of elements is calculated. The class which in majority is found among the first k samples obtained will be the most similar one. Generally the Euclidean distance is used:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{p}(x_{ir} - x_{jr})^2}$$

It is known as “Lazy Learning” because the function is only locally approximated.

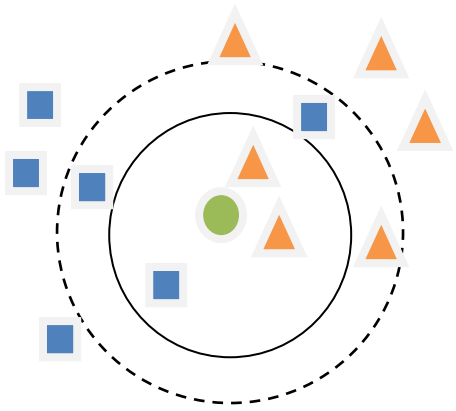

Figure 2- K-NN example

Figure 2 shows an example of how the K-NN method works. The class to be classified is the green circle. In the training phase, the classifier obtains the different classes (red and blue figures) and their position along the space. As we can see, the class which finally will be determined as “closer” to the unknown value would be the orange triangle, as it is the majority.

2.3.2. Hidden Markov Model (HMM)

In this case, we assume that the system that we are classifying is a Markov process. The objective is to determine the unknown or hidden values from the ones that we previously know. It is usually used for temporal pattern recognition such as speech, handwriting, gesture or musical recognition and it can be considered as the simplest dynamic Bayesian network.

The difference between this Markov model and the regular one is that here, the observer is not able to see the state directly, but only the state transition probability along the tokens, which gives some information about the sequence of states. The adjective 'hidden' refers to the state sequence through which the model passes.

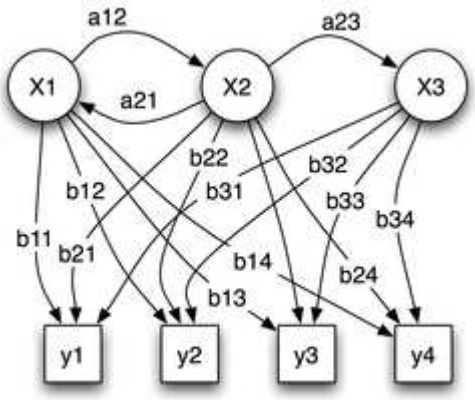

Figure 3- HMM example.

Figure 3 shows the transition between the different states of a Hidden Markov Model. x_i are the hidden states, y_j are the visible outputs, a_{nm} are the transition possibilities and b_{nm} are the output possibilities.

2.3.3. Gaussian Mixture Model (GMM)

It is a probabilistic model represented as a weighted sum of Gaussian component densities used to represent the presence of subpopulations within an overall population without requiring a specific dataset to identify it. It is commonly used with vocal-tract related spectral features in speech recognition systems and handwriting recognition. Despite the fact that the "mixture distributions" can derive the properties of the overall population from those of the subpopulations, this method is used to make statistical inferences about the properties given only from the observation of the pooled population without sub-population-identity information. They use usually unsupervised learning or clustering procedures.

2.3.4. Artificial Neural Networks (ANN)

This is a mathematical model inspired by the biological neural networks in animals. It consists of an interconnected group of artificial nodes (neurons) which processes information using a connectionist approach to computation, adapting its structure on the external or internal information that flows through the network during the learning phase. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

The structure would be the following:

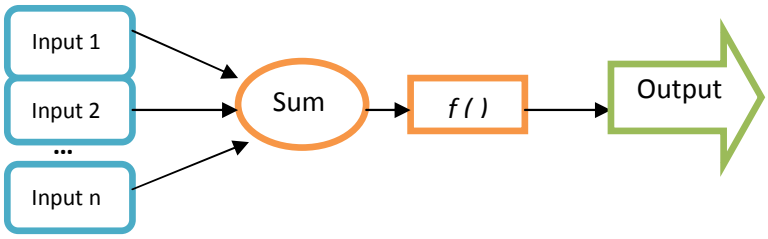

Figure 4- Nodes of the Artificial Neural Network

The objective of this model is to simulate the brain´s performance in a robust and generalized way. Each neuron receives as input the outputs of other neurons through the interconnections. It sums all of them with each determined weight to conform the effective input. Every neuron has a 'state of activation' which can be 0 or 1. If it is 0, means the neuron is not active. Any other value means that the neuron is active. The output of the neuron would be this value. The procedure is simple: the neuron receives the input information from the other neurons and uses it to calculate the output signal that is propagated to the other units.

Figure 5 shows a schema of the neural network:

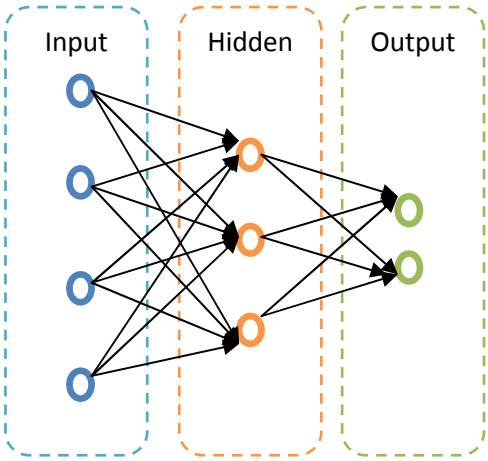

Figure 5- Artificial neural network example.

It is commonly used in real time applications and one of its advantages is that it has error tolerance, for example, to noise, so it is useful when we talk about voice or signal recognition patterns.

2.3.5. Support Vector Machine (SVM)

It is one of the most useful methods using a classifier based on previous examples. They are really close to the neural networks (specifically to the classical multilayer perceptron), but unlike them, the SVM's try to find the decision boundary. One of the advantages of these methods relies on its simplicity, that is, calculating just the geometry of the boundary that separates the different classes.

The Support Vector Machine is based on the induction principle of the *Structural Risk Minimization* (SRM) as a process of inference. More formally, the objective of this method is to construct and select a hypothesis that achieves the largest distance to the nearest training data points of any class, since the larger the margin the lower the generalization error of the

classifier. This will be called *optimal distance hyperplane* because it leaves the cases with one category of the target variable on one side of the plane and cases with other category on the other side of the plane. The figure below presents an overview of the SVM process considering a two dimensions example.

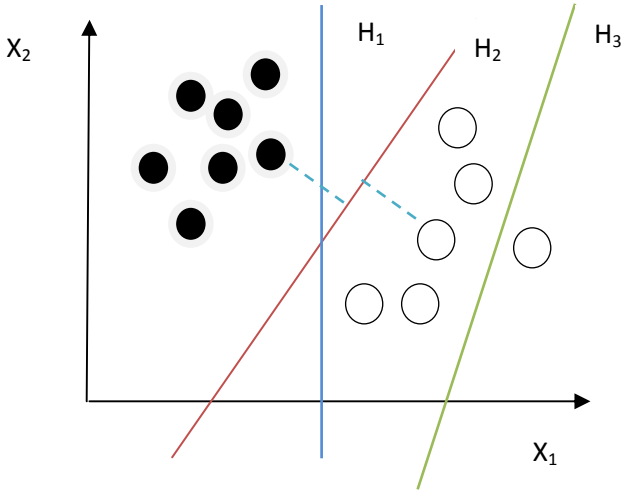

Figure 6- 2 dimensional SVM example

In Figure 6 we can see three possibilities: H3 (green) does not separate the two different classes at all; H1 (blue) separates them, but with the minimum margin; H2 (red) does it so, but with the maximum margin. This one would be the best option. Samples on the margin will be called *support vectors*.

We call *margin* to the distance left between the two parallel dashed lines that separate the space reserved to each class which are drawn as close as possible to the last vectors. An SVM finds the line (two dimensions) or the hyperplane (n-dimensions) that is oriented so that the margin between the support vectors is maximized.

Depending on the given set of training data *D,* there are three different types of scenarios which can be found: Linear, Non linear (non separable data) and Kernel trick.

2.3.5.1. Linear SVM

In this case, we will consider a two dimension scenario like the one shown on Figure 6 above. The set of training samples D is represented as a p-dimensional vector real vector in a R^P space and can belong to two categories or classes y_i {-1,1}.

$$D=\{(x_i ,y_i) \mid x_i \in R^P , y_i \in \{-1, 1\}\}_n \quad , i=1$$

This dataset is considered to be linearly separable. The goal is to find the maximum margin hyperplane that divides the points having y_i= -1 from those having y_i= 1. To draw the hyperplane, we should find the set of points that satisfies the following equation:

$$w \cdot x - b = 0$$

where w is the normal vector to the hyperplane and the parameter $\frac{b}{\|\mathbf{w}\|}$ is the offset of the hyperplane from the origin along the normal vector w. The point is to maximize the distance between the parallel hyperplanes according to the values of w and b. The two possible equations can be:

$$w \cdot x - b = -1$$
$$w \cdot x - b = 1$$

To find the maximum margin between the two hyperplanes and supposing that the training data are linearly separable we have to minimize ||w|| so that the distance is $\frac{2}{\|\mathbf{w}\|}$. Considering that no data points should fall into the margin, the following constraint should be remarked:

$$(w \cdot x_i - b) \geq 1 \quad \text{for } x_i \in \text{first class}$$
$$(w \cdot x_i - b) \leq -1 \quad \text{for } x_i \in \text{second class}$$

Therefore, can be rewritten as:

$$y_i (w \cdot x_i - b) \geq 1 \quad \text{for all } 1 \leq i \leq n$$

where ||w|| should be minimized.

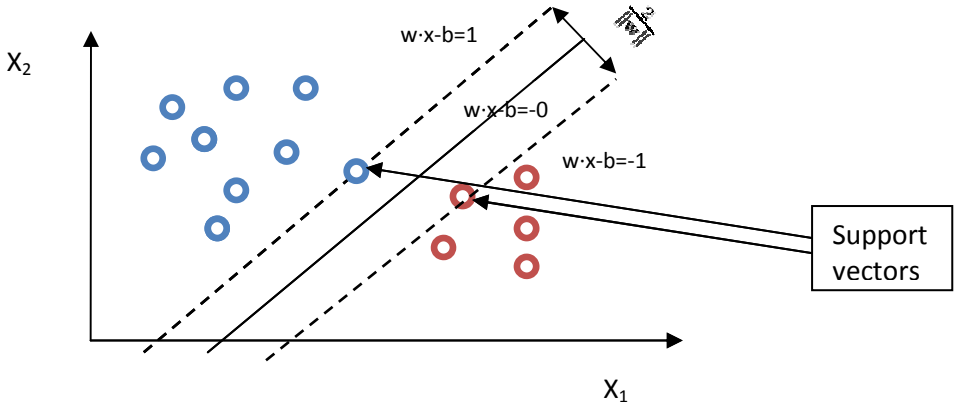

Figure 7- Maximum margin hyperplane for a two dimension SVM

This optimization problem is difficult to solve because it depends on the normal vector to the hyperplane w, what envolves a square root. Fortunately, it is possible to solve this equation by substituting ||w|| by $\frac{1}{2}||w||^2$.

Making use of the non-negative Lagrange multiplier α_i this problem can be solved by standard quadratic programming techniques. The constrained problem can be expressed as:

$$\min_{\mathbf{w},b} \max_{\boldsymbol{\alpha}} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x_i} - b) - 1] \right\}$$

And the solution as a linear combination of the set of training vectors:

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x_i}$$

Where n is the total number of training points and only a few α_i will be greater than zero. The support vectors x_i lie on the margin and satisfy $y_i (w \cdot x_i - b) = 1$ so we can define the offset b as:

$$w \cdot x_i - b = 1/ y_i = y_i \rightarrow b = w \cdot x_i - y_i$$

2.3.5.2. Non Linear SVM: Soft margin

If we consider a set of training data which is not linearly separable, the SVM would not be able to find an hyperplane discriminant enough to split the values. In this case, the optimization problem is relaxed assuming misclassified samples as noisy data. This method is commonly known as *Soft Margin* and splits the examples as cleanly as possible, still maximizing the distance to the nearest cleanly split examples.

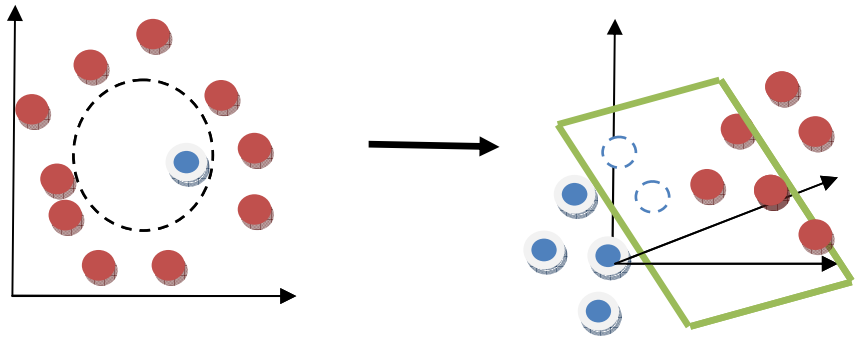

Figure 8 - Non linear transformation example

To measure the degree of misclassification we add the slack variable ξ_i to the equation:

$$y_i (w \cdot x_i - b) \geq 1 - \xi_i \text{ for all } 1 \leq i \leq n$$

This equation is penalized with non-zero ξ_i so the optimization becomes a tradeoff between a large margin and a small error penalty.

Assuming a linear penalty function, the optimization problem can be expressed as:

$$\min_{w,\xi,b} \left\{ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \right\}$$

subject to $\xi_i \geq 0$. The objective of minimizing ||w|| can be solved as in the previous section using Lagrange multipliers:

$$\min_{w,\xi,b} \max_{\alpha,\beta} \left\{ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i [y_i (w \cdot x_i - b) - 1 - \xi_i] - \sum_{i=1}^{n} \beta_i \xi_i \right\}$$

In this equation, we presume $\alpha_i, \beta_i \geq 0$ and the apparition of the constant C as an additional constraint on the Lagrange multipliers is considerate to sum all ξ_i penalizations. The support vectors α_i will solve the equation when $0 < \alpha < C$.

2.3.5.3. Kernel Trick

Unfortunately, the example shown on the Figure 6 is not usual. In daily life, there are usually more than two dimensions, so the separation between classes is not always a straight line and the SVMs have to deal with separating the points with non-linear curves or handling with clusters that cannot be completely separated.

When a linear boundary is not enough to separate the data, a transformation should be done. As shown on Figure 8, sometimes it is easier to transform our feature map to higher dimensions to make it simpler.

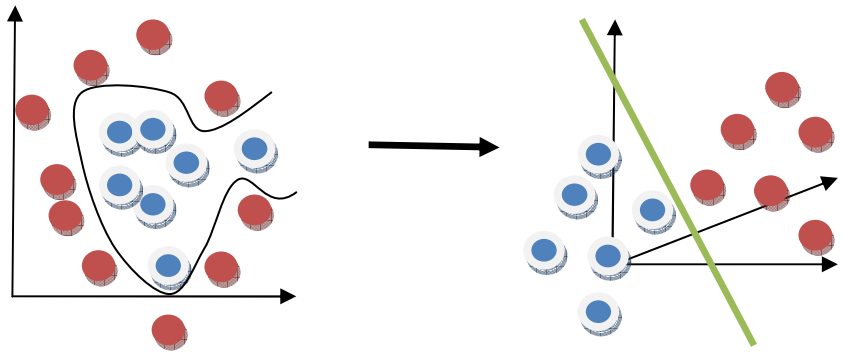

Figure 9- Non-Linear separation of data. Kernel trick.

The *Kernel function* maps the data into a different space where a hyperplane can be used to do the separation rather than using complicated non-linear curves through the data. It is a very powerful method used to perform SVM models. It is referred as the *Kernel trick* to the transformation of the function´s inner vector products to higher dimensions to make them less complex to be solved.

The Kernel functions have been found to work well in for a wide variety of applications. There are many kernel functions to be used but the default and recommended one is the Radial Basis Function (RBF). In this case, the equation that solves the optimization problem would be:

$$K(x_i\ x_j) = \exp(-Y\,||x_i\ x_j||^2) \text{ for } Y > 0$$

2.4. DETECTION OF THE DIFFERENT GENRES

In TV genre we can find a large range of different types of content, such as news, cartoons or talk shows. If we extrapolate the detection of genres to the Internet, the possible range widens. In channels such as YouTube, where the user takes creative capabilities, detection becomes even more difficult.

As seen in the literature, after many studies on the subject, we have concluded that there are certain characteristics that may be useful to discern between the most common genres. According to them, we can distinguish four categories within similar features: Communication, Sports, Information Programs and Entertainment Programs.

2.4.1. Communication

In this section we will focus on commercials. There are some applications specifically designed to recognize the presence of advertisements into broadcast programs. Some possible costumers interested in this type of detection would be electronic companies developing intelligent digital video recorders which detect and remove commercials from the recorded video.

Commercials are easily detected because:

- Duration between 1-6 minutes (whole block of commercials)
- Each one is separated from the others by black frames
- Higher volume audio
- Use of contemporaneous presence of speed and music on audio stream
- Bright colors, quick music and a lot of action
- Lack of TV channel´s logo

2.4.2. Sports

Sport genre includes a great variety of sub-genres, such as football, basketball, tennis, snowboard and many others like the ones used in this work. Although they are all completely different among the others, we can conclude that they have also very similar characteristics. Taking as an example a Football match, some of these features would be:

- Duration limited: 90 minutes
- Selected motion objects: the ball
- Text recognition: the name of a player on the T-shirt
- Face detection: Small percentage of the background refers to the crowd; some sports have more forefront images of the players than others.
- Audio features: high levels of background noise and high speech rate.
- Visual features: most of the background color is uniform.

These features will be explained in deeper detail in section 3.3..

2.4.3. Information programs

In this section we are going to talk about news and weather forecasts. Both genres are well known for their regular structure so that they can be easily identified by the viewer. Other features are:

a) News:
   - Duration between 10 and 30 minutes
   - Regular structure: Anchor person always in the same position
   - Two types of shots:
     - Anchor person shots: repetitive and spread along the screen
     - Report shots: reporter speaks over external report
   - Static background and few movement
   - Few different colors and structural changes
   - Mainly face shots
   - Low speech rate

b) Weather forecast:
   - Duration between 1 and 3 minutes
   - Mainly satellite images (background)
   - Face position is always on the left and right side

- Speech always about what happens in the video
- Few repetitive shots
- Possible detection of clouds, suns, colored capital letters…

It should be remarked that some of the characteristics seen in News can also be found in genres like Talk shows and interviews. On the other side, the distribution of the frames and the synthetic style of the backgrounds are typical just of the weather forecasts.

2.4.4. Entertainment programs

Under entertainment programs we can include such a big variety of genres that we will have to focus just on the most famous ones to generalize. These are:

a) Cartoons:
   - Variety of characteristics depending on the age of the targeted audience and the date in which was created
   - High variable of semantic content
   - Bright colors, what means, saturated histograms
   - High speech rate
   - Usually, high motion

b) Music videos:
   - Rates from live concerts to music clips
   - Each clip is one single song so duration is between 2 and 6 minutes
   - Important audio features
   - Bright colors and high motion

c) Talk shows:
   - Broad variety of programs: game shows, informative shows, sport shows, simulated legal encounters, interviews…
   - Many have public participation and live telephone calls
   - Face to face discussions: face distribution is nearly constant
   - Duration between 20 minutes and 2 hours
   - Many repetitive shots
   - Low average speech rate

2.4. DISCRIMINATION OF SPORTS AMONG OTHER GENRES

Among all possible genres discussed in the previous section, for the realization of this work we have decided to focus the attention on the detection of sports. The main reason is that the features inherited from the news summarization system [1] were focused on detecting color, face parameters or movement characteristics, which apparently seemed to suit sports classification efficiently.

Nevertheless, as discussed in the section destined to explain the future work (Section 5.2), the idea of this framework is to develop a methodology based on the classification of sport video summaries that can be afterwards extended as a workflow to be able to build new systems to classify many other types of videos. This will require the adaptation of the feature selection depending on the genre that wants to be detected.

2.5. DISCUSSION

Based on the previous work, we can conclude that the automatic classification of video can be very intricate due to its awkwardness of defining the concept of genre in an objective way in order to be recognized worldwide. Analytically, it results also difficult to build a sufficiently differentiated taxonomy. This is why most projects in this area have focused on classifying just a single genre [15, 16, 17, 18, 19] or a limited number of classes distinguishable enough [3, 4, 5, 7, 8, 15].

Given the conditions discussed in previous sections, the algorithm that best fits the automatic video classification should be simple, so there are no large computational costs in case of classifying huge amounts of video, and should also be highly applicable because the gender of those videos can be variable. The two options that best suit our purpose would be Support Vector Machine (SVM) and Artificial Neural Networks (ANN), but we are just going to focus on the first one.

The algorithm chosen for this project is the Support Vector Machine (SVM), mainly for its simplicity and efficiency. This technique is based on learning by examples, so it fits perfectly for our purpose. Unlike the neural networks that try to build a model a posteriori, the SVM's key advantage is that it concerns just finding geometrically the decision boundaries, which results much easier in many cases. In turn, it should be noted that such algorithms are based on the classical method of statistical pattern recognition, which means dealing with the drawback that its efficiency depends on the separability of the features used to represent the set of multimedia training data. This problem can be solved using algorithms artificial neural networks as they are robust enough to handle noisy data. The possibility to expand this classification procedure in the number of detected genres makes the SVM method the best choice since it

CHAPTER 3. DESIGN AND IMPLEMENTATION OF AN AUTOMATIC VIDEO CLASSIFIER

3.1. Introduction

The work presented in this document starts from a base system developed to provide an efficient, automatic and on-line method to summarize news bulletins [1]. The result of this work is a video abstract that for each story found during the news bulletin generates a visual composition combining the anchorperson introduction and a video skim of the visual segments of the story. This system is composed of 4 phases (Analysis, Classification, Skimming and Composition) from which we are going to use just the first two ones, ignoring the ones related to video summarization.

To understand easily how the system works, a block diagram of the proposed framework is depicted on Figure 10.

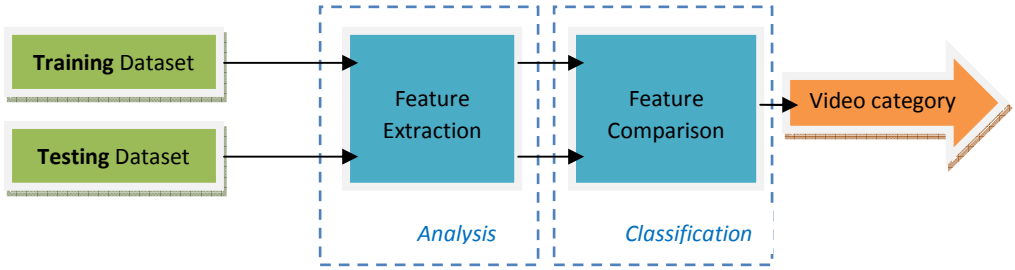

Figure 10-Block diagram o the proposed framework

As shown on Figure 10 our system is divided in two phases: Analysis and Classification.

- *Analysis:* It is at this stage where the input videos are divided in small segments and, for each one, low-level features are extracted. These features, such as MPEG-7 Color Layout, frame differences, color analysis and face detection (see section 3.4.) will be used for the next step: classification.

- *Classification*: Each testing video segment received has previously been annotated in the analysis stage, so it is classified based on the information provided by a set of SVMs, independently trained with the training datasets, for the different possible shot categories.

According to our framework, sport video classification, we can distinguish 8 different categories: *Football, Basketball, Tennis, Beach volley, Snowboard, Water sports, Formula one* and *Cycling*. It is important to take into account that the system is designed to work with any kind of video without assumptions about their length or shot composition.

In the remaining sections, the implementation details of each individual component are provided. First, section 3.2. overviews the *Architecture of the system*, outlining each module and the software used in each phase of the simulation. Next, section 3.3. *Database selection and transformation* focuses on the type of database used. To end with, sections 3.4. and 3.5., *Proposed feature sets* and *Classification phases*, describing, respectively, the features selected for our purpose and the different phases up to the optimal classification.

ID Number	Category
1	Football
2	Basketball
3	Tennis
4	Beach Volley
5	Snowboard
6	Water sports
7	Formula 1
8	Cycling

Table 1- Category number identification

3.2. Architecture of the system

As mentioned in the preceding section, our framework is based on a previous system. The architecture of our system aims to study the feasibility of an automatic sports video classifier and it is clear that, according to the inherited architecture, some modifications are needed. The database selection and the features addition are the most significant changes.

The following figure illustrates the modules of the new architecture of the system. Note that the subindex n refers to the category of each video, taking the value 1 to 8 respectively with the categories named on section 3.1..

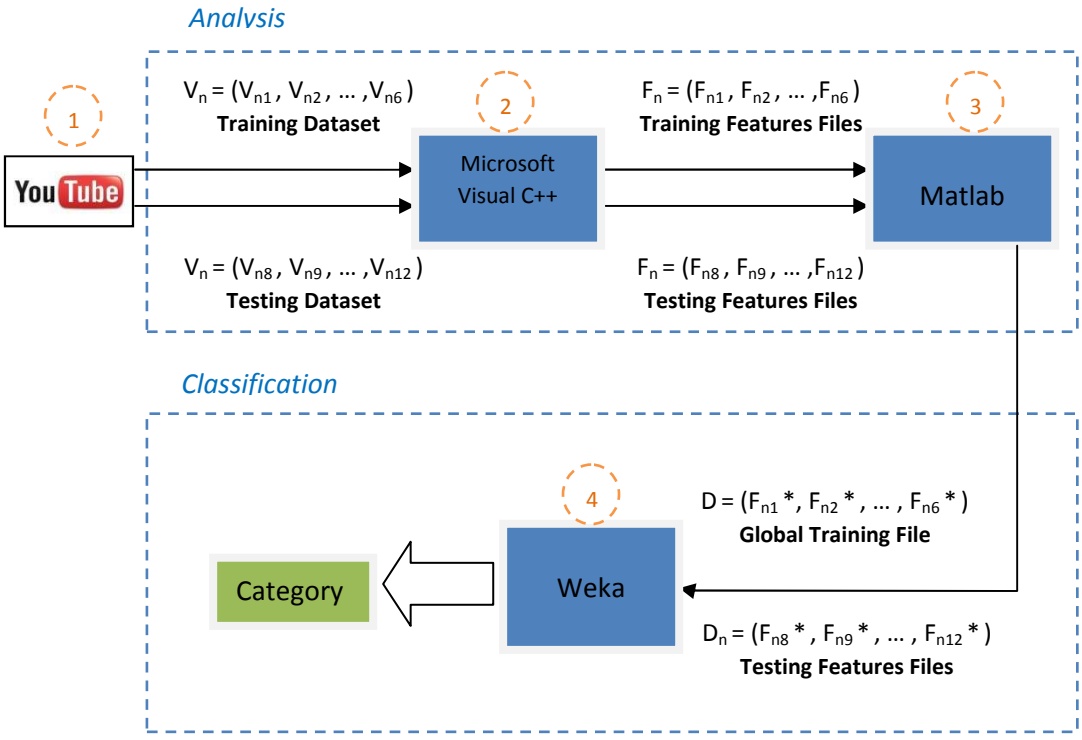

Figure 11- Block Diagram of the architecture of the system

As depicted on Figure 11 , we can divide the whole schema into 4 phases:

1. *Database creation:* Videos have been carefully selected from Youtube to compound a database robust enough to have feasible results. This part will be explained in detail in section 3.3..

2. *Feature extraction of each video:* Once our database has been transformed into mpeg video format so that it is compatible with the code, each video is used independently as an input to the algorithm developed with Microsoft Visual C++ 2010 to extract the selected features. This way, we obtain one feature file for each video. The description of the features can be seen in section 3.4..

3. *Execution files creation:* In this stage, we use Matlab to generate the definitive files for the classification. For this purpose, each input file has to be labeled with the number of video and the number of category to be afterwards identified. This is the reason for the asterisk. As the figure above illustrates, there is only one global training file created from the union of every single training file. On the other hand, the testing files will remain always individual (see section 3.5.).

4. *Classification:* For the classification we are going to use Weka [20]. We introduce both files, training and testing, and the machine, comparing the values of the testing features to the ones on the training file decides which category suits best the testing video. This is also explained in detail in section 3.5..

The hardware platform used for the execution of this work has been an Intel Core 2 Duo @2.26GHz with 4GB of RAM.

3.3. Database selection and transformation

The first step to create a system is the selection of an appropriate corpus of videos. As mentioned in previous sections, the database used to train and test the framework has been selected from Youtube due to its accessibility and the broad variety of contents. Another reason is that one of the possible applications of our system can perfectly be the automatic classification of video on this channel.

Youtube offers many different categories labeled by the users and intends to maintain reliable on the classification despite being an open source where anyone can upload videos and subjectively categorize them. This tool permits the download of the videos and has several applications in different programming languages to be integrated in other systems. One of the characteristics of the videos loaded in Youtube is that each one has an information sheet with complementary parameters about the video, such as author, date or for example the category.

According to the purpose of this work, we have focused on selecting just sport videos, concretely 12 for each category. One of the drawbacks of a database composed of videos extracted from Youtube is that the quality and the structure of the content is not homogeneous. Before the start of the simulation, some previous decisions are taken as explained at the beginning of chapter 4; for example, the number of videos destined to the training and to the testing phase or the type of classification, binary or multiclass.

It must be noted that the downloaded videos that form the database have been previously converted to a common format MPEG so that the code in Microsoft Visual C++ 2010 would accept them.

3.4. Proposed feature sets

In this section, we overview the features used for the classification of sport videos. We distinguish two kinds of features: the old ones, inherited from the base system, and the new

ones, those which have been added to improve the performance of the framework. Both are explained in the following sections.

3.4.1. Inherited features

Most of the features used in this system have been inherited from the base system [1]. We can group the features into six modules:

a) *Face Detection*: Based on Haar features [21, 22], the OpenCV library[1] provides a very fast method for arbitrary object detection. One of the main objectives of the base system was the detection of the anchorperson, so this feature is mainly focused on frontal face detection. The average number, size and coordinates of detected faces as well as the variance of such features are also calculated. Both can also be used to differentiate between sports.

b) Color Variety: This module differentiates between natural and synthetically generated images. It measures the number of representative colors in an image using the Y, U and V channels histograms are calculated.

c) Frame differences: This kind of features try to make an estimation of the video activity by calculating the average variation between consecutive frames. Making use of the Color Layout Descriptor extraction, an 8x8 thumbnail image is generated for each decoded frame. As shown on Figure 12.(a) , five different activity areas have been defined in order to distinguish between different types of activity, for example, local or global motion patterns.

d) Shot variation: In order to provide a different activity measure to that obtained with the thumbnail subtraction, an average segment variation measure is obtained. The Color Layout difference is calculated every 3 frames and then is averaged.

e) DCT Coefficients Energy: The Discrete Cosine Transform coefficients, DCT, make up the Color Layout descriptor in every 8x8 thumbnail. These pre-calculated parameters measure the energy distribution of the images, characterizing the ones with smooth or abrupt changes. In this case, the descriptor coefficients have been divided into four areas which are added up and averaged to obtain four frequency measures. Figure 12.(b)

f) Image Intensity: The mean and variance of the intensity of each frame are calculated and averaged for each video segment. This is used to detect constant and controlled illumination conditions, useful, for example, to classify indoor sports.

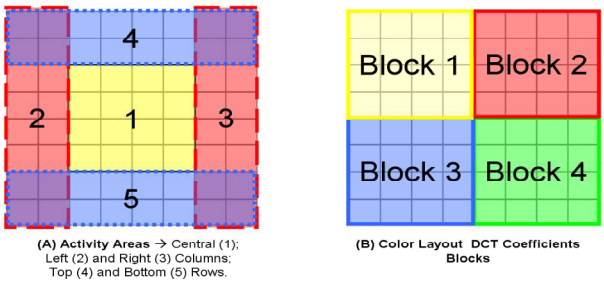

Figure 12- (a). Frame Block Variation Areas. (b). DCT Coefficients Blocks

[1] http://sourceforge.net/projects/opencvlibrary/

It is important to remember that these descriptors were chosen for the on-line news bulletin summarization, and despite most of them adapted correctly to our new purpose, during the development of this work several studies have been done to measure their feasibility. For example, it has been decided from the obtained results which descriptors perform best and which should be deleted or even the order in which they should be applied. The results of these procedures can be found in Chapter 4. To improve the performance of the system according to sport video classification, we have added two new descriptors that will be explained in the following section.

An overview of the features in this work, according to the module division previously explained can be seen on Table 2.

Face detection	Color variety	Shot variation	Frame differences	DCT Coefficients energy	Image intensity
faceRate faceX faceY faceRadious faceVarX faceVarY faceVarR	colorRate cY cCb cCr	averageThumbnail matrix 0 1 1 0 2 2 0 1 6 0 2 5	Variation maxVariation startEndVar Comparisons: - arriba/total - abajo/total - izquierda/total - derecha/total - central/total	CLEnergy [Q_1, Q_2, Q_3, Q_4]	Intensity maxIntensity varIntensity

Table 2- Overview of the visual features used

3.4.2. Added features

As mentioned in the previous section, some features have been added in order to improve the performance of the system. The main reason of this addition is that the inherited characteristics were selected to summarize on-line news bulletins and, despite the efficiency obtained for this purpose, they do not always offer the expected results on sport video classification. During the post-process of the framework the need of new features became imperative and it was decided that, according to the content of the videos, descriptors related with color would be the ones which would perform best. Making use of previous work [23, 24], it was determined that the new features would be *Color Layout* and *Dominant Color.*

a) *Color Layout:* This descriptor was used in [23] to make MPEG-7 descriptions on a set of images and then measure the similarity between different images based on the descriptors generated. The main characteristic on which these descriptors are based is Color Layout. The extraction is depicted in Figure 13:

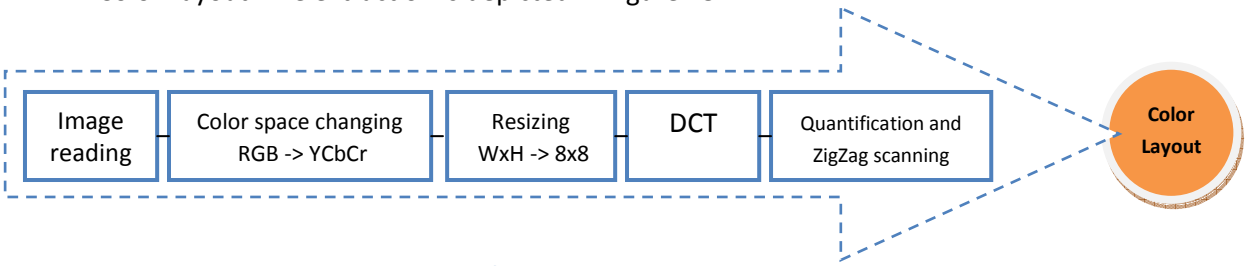

Figure 13- ColorLayout extraction process

To generate the MPEG-7 description the information of ColorLayout will be used as visual descriptor.

To calculate the phases "color space changing" and "resizing" we have used the appropriate functions of OpenCV. The result of this extraction are 64 coefficients resulting from DCT+Quantification for each plane of image (Y,Cb,Cr) and 6 and 5 bits for the DC and AC coefficients respectively.

b) *Dominant Color:* This descriptor specifies a set of dominant colors in any arbitrary shaped region. It targets content-based retrieval for color, either for the whole image or for any arbitrary shaped region, either rectangular or irregular. It is important to know that the color quantization descriptor does not support non-uniform colors.
The Dominant Color extraction algorithm takes as input a set of pixel color values from the RGB color space and quantizes the color vector in the image based on the Generalized Lloyd algorithm (GLA). The dominant colors are extracted as a result of successive divisions of the color clusters with the GLA algorithm in between each division and then merging of the color clusters. (see [24])

Both features have been chosen for their characteristics related to color extraction. The ground of this decision is that it has been studied that sport video categories can be differenced mainly for their background color, for example, to differentiate football and snowboarding, you just have to look at the dominant color, green or white respectively. These two features can be added as part of Table 2.

To be considered for future work, motion activity can be added too in order to obtain good results according to sport video classification. The principal drawback is the complexity of the implementation of this feature.

3.5. Classification phases

Automatic video classification is not an easy task. First, it must be decided which features must be extracted so as the categories that will be detected during classification. Then, once we have the extracted features, we proceed to see whether they are suitable for the selection of the categories that have been chosen, differentiating efficiently between them. It is because of this difficulty that we have decided to use a broadly known classifier: the Support Vector Machine (SVM), which has been proven to provide a good performance in different classification problems [25].

In this work, we have considered two different types of classifications: binary and multiclass. The first one, binary classification, consists on the detection of just one category at the time, this means, we introduce a video fragment to 8 classifiers, one per category, and depending on the similarity of the features extracted, we obtain as output an 8 bits vector. If one classifier considers that the video fragment belongs to its category, the output will be 1; if not, the output will be 0. This is shown on the step 1 of Figure 14. The problem begins when more than one classifier sends an active response. Although an SVM with individual binary classifiers usually provides a very good starting point, the final decision about which category does a frame belong to is not always straightforward. When we have a multiple positive situation, we observed that a multiclass global classification would work better.

Our simulation is not exactly as the one depicted on Figure 14. The SVM used for this work is a multiclass classifier and the results are subjectively evaluated by the percentage of instances correctly classified. The decisions of implementation taken before the final simulation can be seen in section 4.5.1..

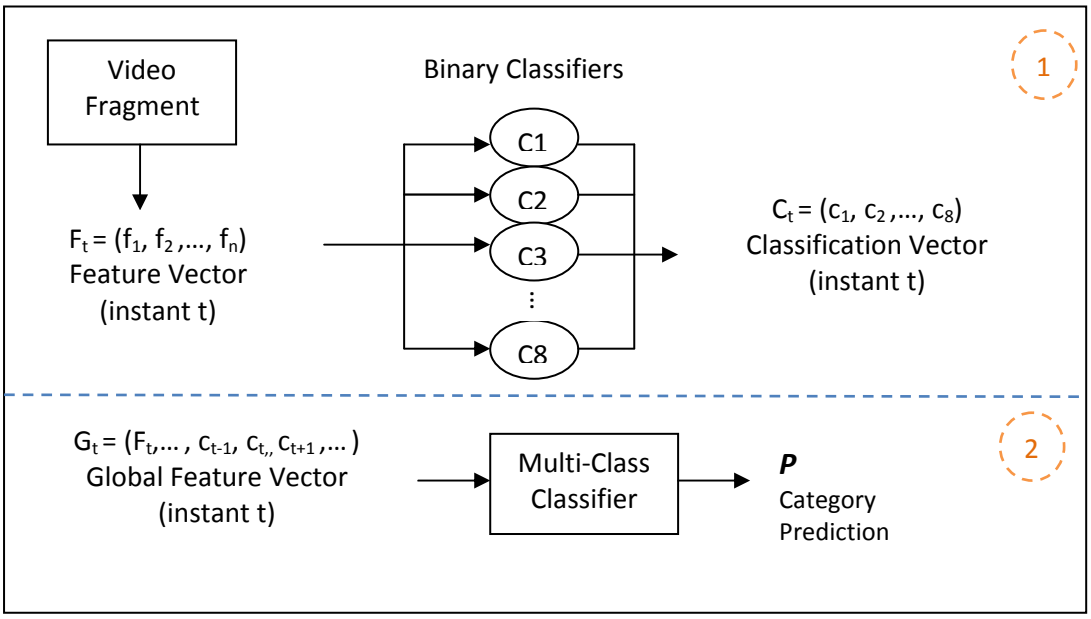

Figure 14- Multi-class Classification steps

3.5.1. Training and Testing

The procedure is divided into two phases: training and testing. As explained in section 3.3. , we have gathered the total amount of twelve videos for each of the eight categories. From these twelve videos, half will be used for the training and half for the testing phase. It must be said that both, videos and documents, used for the training and testing phases have been stored separately from the beginning. This way, there is no possibility to mix the parameters from both phases what would vary seriously the final results.

For the training process, the procedure is the following: each of the six videos selected for the eight categories would be sampled every 25 seconds to form a vector of frames per video. Once the features are extracted, each of the frames will be manually annotated with the label [frame, video, category, feature_vector]. The idea is to collect all this vectors into a single document with which the classifier will be fed during the training process and builds up the boundaries between the different category spaces. In this case, as commented in section 3.2. we have chosen Weka as our SVM classifier. It is important to remark that every training document has previously passed through a supervised filter by instances called "*SpreadSubsample*" that produces a random sub sampling of a dataset to make sure that there is a balance between all the categories. The spread distribution must be 1. This is done to give equal possibilities to all the categories; if not, the category with more samples would be the most common one and by the way more likely to be chosen.

Once our SVM is trained, we proceed with the testing. The modus operandi is the same as in the training phase, but for the database. Every video is labeled and introduced in Weka. These videos do not need to be filtered. We introduce in Weka the filtered training file and we select the function "*libSVM*" in the classify tab. This classifier must be normalized. The testing file is browsed in the supplied test set to start the classification. From the output data we focus on: instances correctly or incorrectly classified and their percentages, a detailed "Accuracy By Class" and a "Confusion Matrix" (see section 4.3.). All this information helps us to analyze if the classification is working correctly or if there is any kind of problem with the training data, the content of the videos or the features selected.

An important point that must be tested out is the detection of burst errors. It can be checked selecting on the menu “More Options” the box “Output predictions “. This way, we can check frame by frame if the classifier has guessed correctly the category of the video according to its features or not. It can also be useful to measure the consistency in the category of consecutive video frames. If it happens that a group of consecutive frames are incorrectly classified, we consider it as a “burst of errors”, this is, that the sequence of the video contains images that are not similar to the category to which the video belongs to. Although this is not easy to detect, to reduce possible classification mistakes caused by ambiguous videos which do not seem to belong to a specific category, we apply a “Median Filter” to the output vector to determine the most common category. This procedure is explained in detail in Chapter 4.

3.5.3. Evaluation

The most important part of this project has been the evaluation of the results. It can be said that this system has been under constant evolution, always based on an improvement in the overall efficiency of the classification. The different parts of the pre- and post-process and the results of each one can be seen in Chapter 4.

CHAPTER 4. EXPERIMENTAL EVALUATION

4.1. Introduction

This chapter makes a detailed overview of the successive phases that have been followed during the development of this system. At first, there are given some specifications about the dataset used (section 4.2.), the performance measures taken into account (section 4.3.) and the experimental settings of the framework (section 4.4.). Next, section 4.5. illustrates the experimental results of the proposed classification system. It is divided into five subsections, each one according to one of the experimental phases of the process.

4.2. The experimental dataset

The selection of each experimental database depends on the aim of the specific research work under investigation. The objective of this work was to classify randomly selected sport video summaries extracted directly from Youtube with variable lengths and structures. The heterogeneous nature of the videos, although they belong to the same category, affects right away to the obtained results. If the videos followed a pattern or had been edited and produced by the same user/TV channel, this would have been an advantage for the classification and it would have improved the final results. Besides, videos have been selected to intend to ensure genre representativeness without turning away from reality. The ability of our system has been tested in classifying the eight different sport categories already mentioned: *football, basketball, tennis, beach volleyball, snowboarding, water sports, formula one and cycling.*

As explained in section 2.4. each of these particular sport categories has its own characteristics to be distinguished among the others too. For example, football videos will probably have during the whole duration lots of green background frames. On the other hand, snowboard will have white background and formula one grey background. This way, the dominant color of each frame would be relevant information to be taken into account. So as other features related to color (color rate, luminance or color layout energy), others related to face detection (face rate and face distribution) are also distinctive of each category.

4.3. Performance measures

For the evaluation phase, we have made several performance measures with a suite of machine learning software written in Java called Weka. This application is responsible for the training and testing classification phases. When we introduce the training and the testing files to start the simulation, we must know what kind of outputs we are going to obtain in order to understand the results. From the output file we are focusing on the following data:

a) *Instances correctly or incorrectly classified*: In this section we study the percentage of hits we have obtained with the classification. The number of correctly classified instances corresponds to the number of frames for which the classifier has guessed correctly the category. It is important to ensure that this value is not too low in order to achieve good final results.
.

b) *Accuracy by Class:* In this case we are going to focus just on two values: Precision and Recall. In pattern recognition and information retrieval, Precision is the fraction of retrieved instances that are relevant, while Recall is the fraction of relevant instances that are retrieved[2]. In this case, we should check that the percentage of these values remain medium and balanced. A high recall means low precision: although you haven't missed anything you may probably have a lot of useless results to sort through. On the other hand, following the same procedure, high precision means low recall: if everything returned was a relevant result, you might not have found all the relevant items.

c) *Confusion Matrix:* This is the most important information in the output file. This matching matrix shows how the classifier distributes, according to the feature values, the global amount of frames in the eight categories. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. It is very useful to detect if the system is confusing two or more classes, for example, when a video from category A has a big quantity of frames similar to another category or when, during the training, the dataset has not been balanced (A: 9 samples, B: 1 sample) and the classifier biases towards the majority class. For this reason, we use during the training the "SpreadSubsample" filter.

This information is used during the whole development of the system to check if the simulations are working correctly. It has helped in many cases to detect problems such as not useful training data, not well defined content of videos or errors related to the features used.

4.4. Experimental settings

For the experimental part, we have used a prototype developed for the summarization of on-line news bulletins [1]. The feature extraction was developed by the author in C++ as programming language making use of the libraries provided by OpenCV. We can highlight the library used to obtain the face detection features (Haar features) and the Color Layout descriptor. The gathering of the data and the generation of the files has been done with Matlab R2010a for Windows 7. The classification was performed on Weka (Waikato Environment for Knowledge Analysis) where the SVM is already implemented.

As it would be explained in section 5.1., several studies have been made to decide that from the global database, six out of ten videos will be used for the training phase and the rest for the testing. This parameter is called "split percentage" and the system has been designed so that this value can be modified by the user. At the end of the process, it is set that all the categories must have the same number of videos (twelve, six for training and six for testing) and it will be decreased from 60% to 50%. The classification has been done with a multiclass classifier.

4.5. Experimental results

This section illustrates the experimental results of the proposed classification system. It analyzes the individual classification effectiveness of each descriptor and the combined performance of the ensemble system.

[2] http://en.wikipedia.org/wiki/Precision_and_recall

4.5.1. Previous analysis

Before the simulation, some design decisions must be taken. This section concerns some important concepts that have been considered during the pre- and post-processing of the frameworks development. This system has undergone numerous modifications for its optimization, as it will be explained in the following sections.

As studied in the literature, there are two possibilities for the classification: binary and multiclass classification. The main difference between them is that a binary classifier detects one class among the rest, and the output of the classifier is "1" if it guesses correctly or "0" if not. In the multiclass classifier the output is a weighting among all classes, that is, what percentage of the video has similar characteristics to each of the classes. The highest percentage corresponds to the predicted final category. This section explains the previous stages that have been studied for two reasons: to check that the system works correctly and to decide whether it is better binary or multiclass classification for our purpose.

To get started with the simulation the first thing to have ready is the database. It is composed of 10 videos per category (later expanded to 12) randomly selected from Youtube. The first decision to be made is the number of videos that are going to be used for the training phase and the ones for the testing phase. This "split percentage" represents specifically the amount of videos used for the training phase. For the binary case, we run the simulation for four possible percentages: 20%, 40%, 60% and 80% and the conclusion is that, on account of the percentage of correct instances, the 60% is the most effective one. The results for the binary classification can be found in the Appendix I.

Once we know the value of the Split percentage for the binary classifier, we analyze it for the multiclass classifier. The method used for it is called "False Label". The procedure consists on "tricking" the machine by keeping the training files the same, but introducing the same video input in each simulation with a different category label. For example, we choose a football video (label =1) and we simulate it as it was a basketball video (changing the label=2), next as if it was tennis (label=3) and so on with the eight categories. The same procedure as in the binary case is applied for the multiclass classifier, but this time basing the results on the confusion matrixes. It happens that in this case we obtain the same ratio for the 40%, 60% and 80%, so we perform with the 50% and the 100% too. The best results correspond to the 100%, but this is quite tricky because we use the same videos for training and testing so the results are not fair. Following, we find the results for the 50% which are better than any of the other three percentages. This is the main reason for the extension from ten to twelve videos (six for training and six for testing) in the multiclass classifier. To check the multiclass classification results see Appendix II.B. .

This test helps also to validate whether the video contents are easily differentiable or if they have similar characteristics between different categories. Thus, we can see if there are classes that the classifier may confuse and, in that case, try to find better descriptors to improve the classification. The application of this method to the binary classifier was even much easier. We maintain the same testing file with the active label during the whole procedure and we change in each simulation the active category in the training file. The results can be checked in Appendix II.A.. From the results we see that the categories which are better differentiated are 5 (snowboard) and 7 (Formula One).

The next step was to decide if the classification would be binary or multiclass, so we proceed with the testing phase. The results demonstrated that the system worked really good for some categories and really bad for others, in both cases binary and multiclass. The reason lies in the

features which adapt better to the content of some specific videos than others. Anyway, the results conclude that the multiclass classifier performs better than the binary one, so from this point to the end of the work the classification will always be multiclass.

One of the possible causes of obtaining lower results than expected might have been due to an inconsistence on the similarity between the training videos and the testing ones. The videos had randomly been chosen from the Internet, and although they belong to the same category, they might not have similar characteristics. For instance, if we classified tennis videos, the color descriptors will not be the same depending when watching a Roland Garros video than when watching a Wimbledon videos. The second tournament may be detected as football either because of the green background. It can happen that a video sequence has a number of consecutive frames that have been classified as another category. These types of errors are called "burst of errors". To detect the existence of videos with these characteristics, we conducted a thorough study using a "Middle Window Filter". In this case the filtering is done with a window of N=5 frames, but this value can be changed. The algorithm selects five consecutive frames and evaluates the majority category. If there is no majority, it outputs zero instead. The procedure is repeated until the video ends and finally calculates the total majority category of the video. The following example illustrates the procedure:

With a window of N=3, we obtain the following sequence of frames:

…	111	114	173	212	…
	1	1	0	2	

This way, category 1 represents the 50%, category 2 the 25% and the remaining 25% represents an undefined category. The video definitive category would be 1. If this video belonged to any category distinct from 1, it would not be suitable to obtain good results because the content of the video would be more similar to football than to any other category. It should have been deleted and substituted by another one. The results of the “burst of errors” analysis can be found in Appendix III.

At this point we have already checked some of the most important issues to ensure that the simulation is relatively fair. The following sections describe the different phases passed for the optimization of the system.

4.5.2. Individual features performance

Once the system has already been tested in the implementation details, we proceed with the simulation. For this phase, the number of videos has already been extended to twelve (six for training and six for testing). This part of the work consists on trying each video with every single block of descriptors (see Table X) individually and make an analysis of which features perform better with each category. The results can be found in Appendix IV. The nomenclature used for the descriptors in these tables can be found in the following legend:

The main objective of this simulation is to evaluate the features and the videos individually in order to specify the order in which they should be applied. These results will be analyzed in the next section. It is logical to think that if we apply first the descriptors with better results and last the ones which work worst in an accumulative way, the percentage of correctly classified instances will decrease more slowly than if we applied a random order.

Number	Descriptor block
1	Face Detection
2	Color Variety
3	Frame Differences
4	Shot Variation
5	DCT Coefficients Energy
6	Image Intensity

Table 3 - Legend of descriptors

4.5.3. Selection of features order

To decide the order in which the features shown on Table 3 will be applied, we can follow two possible discussions: Subjective Evaluation and Objective Evaluation.

a) *Subjective Evaluation:* Concerns an order decided by the authors of the simulation according to their opinion about which descriptors should intuitively perform better with the categories that are being classified. Due to the choice of classifying sports, it is logical to think that the features that will obtain better results are the ones concerning color characteristics, as explained in previous sections related mainly with the color of the background. Next, following this line, we would focus on the intensity of the images, useful between indoor and outdoor sports. Motion activity is also important when talking about sports. It can also be measured comparing the differences between consecutive frames because it is not the same football as, for example, golf, although the background of both is green. Last, because of its lack of relevance in sports, would be face detection features. So, the order selected by the authors would be:

$$\{5 \rightarrow 2 \rightarrow 6 \rightarrow 3 \rightarrow 4 \rightarrow 1\}.$$

This is the same analysis that has been followed for the addition of features.

b) *Objective Evaluation:* In this case, we have two options: the order given by Weka and the order that has been calculated from the results in Appendix IV. To start with, we have applied the function "Attribute Selection" from Weka to our training files to distinguish which features are statistically useful and which ones are not. The suggested order was:

$$\{5 \rightarrow 3 \rightarrow 2 \rightarrow 4 \rightarrow 6 \rightarrow 1\}.$$

From this order we can see that Weka agrees with our order giving more importance to color and motion activity characteristics and less importance to face detection. It is image intensity that feels not so relevant as in our opinion.

The last option that we have studied is the individual descriptors analysis, seen in the Appendix IV. An organization by class has been made to order video by video which features have obtained the best results and which ones are not relevant at all. Then, all this information is gathered in tables to see which feature primes as the best and the worst in the majority of the cases. This analysis can be found in Appendix V and it has been resumed in table 4. The objective order based on the previous results is:

$$\{2 \rightarrow 3 \rightarrow 6 \rightarrow 4 \rightarrow 5 \rightarrow 1\}.$$

Although it might not be the best according to logic, the order given by the results in Appendix IV is the fairest one. This is the main reason why we chose this order and not the other ones. Note that this analysis would no longer be valid from the moment a different database is used or other categories are classified. Besides, the general procedure could perfectly still be used but the results would be different.

Category:	1	2	3	4	5	6	7	8	Final
Best	6	2	4	6	2	2	1	4	2
↓	3	6	6	3	4	3	2	2	3
↓	5	4	3	4,5	6	4	3	6	6
↓	4	3	5	2	3	5	5	3	4
↓	2	1,5	1	1	5	6	4	1,5	5
Worst	1		2		1	1	6		1

Table 4 - Analysis and Selection of the feature order application

4.5.4. Combined features performance

According to the analysis explained in the previous section, we proceed to repeat the simulation accumulating the features in the selected order: {2 → 3 → 6 → 4 → 5 → 1}. Each video will be simulated on the first place with the descriptor 2 individually. Then, we simulate with the descriptors 2 and 3 and so on until we use all of them. The results can be found in Appendix VI.

One of the conclusions obtained from the results is that the percentage of correctly detected instances has increased compared to the one obtained with the individual descriptor analysis. This means that the chosen order improves the overall performance of the system. The only point that must be taken into consideration is the elimination of the Face Detection descriptor. As it can be seen in the results, in most of the classes it does not contribute to improve the number of correctly detected instances. Moreover, it usually leads the classifier to confusion. For future work, this descriptor should be either deleted or designed to be more useful according to the content of the videos.

Surprisingly, the descriptor number 5 (DCT Coefficients Energy) does not work as good as expected. From a subjective point of view, color characteristics should be relevant to distinguish between sport categories, so, to compensate the poor accuracy of this descriptor, two more color descriptors are added: Color Layout and Dominant color. These will be explained in section 4.5.5..

If we look at the first tests we did (section 4.5.1.) the classes that seemed to fit better with this classifier were classes 5 and 7. Analyzing the results obtained in this section, we observe that these two classes can be detected just with descriptors 2 and 3. Therefore, it can be concluded that for these two classes a descriptor of color and a descriptor of motion activity are sufficient to classify them correctly.

In order to check if the accumulation of features in the selected order really improves the final percentage of correctly classified instances, in Appendix VII we can find some tables in which we study the improvement and declining of the performance. As we can see in the results, it is usual that in some videos the addition of features decreases the percentage, although the

video is overall well classified. This is not fair to blame only the features of being imprecise. Most of the times, the improvement or declining of the results depends on the content of the video that does not conform specifically to these features. For example, if we test the system with a video that has no faces in most of its frames, and we study how it works with Face Detection classification, probably the results will not be very good.

As a general conclusion it can be said that although the results are not bad, they can be improved. For that reason, next section explains the improvements caused on the classification by the addition of new features.

4.5.5. Addition of features

The simulation using inherited features in the selected order achieved good results, but the need of new features that really adapted to the content of the videos became imperative to accomplish the optimization of the system. According to the subjective evaluation, these videos were chosen mainly for their color characteristics, so the new features should be related with color. We decided to use some feature extractors that had previously been used for this purpose and had demonstrated to obtain very good results. This section explains the implementation of two descriptors: Color Layout and Dominant Color.

a) *Dominant Color:* As it was explained in chapter 3, this descriptor specifies a set of dominant colors in any arbitrary shaped region. We make use of some previous code in order to extract this feature. The structure in which the output per frame is obtained would be, for example:

```
RESULTS
Size:                    3
SpatialCoherency:        0
Percentage: 8
Values: 0 0 0
---------------------------------------
Percentage: 3
Values: 9 10 7
---------------------------------------
Percentage: 4
Values: 8 12 5
---------------------------------------
Percentage: 0
Values: 16 12 8
---------------------------------------
```

Figure 15 - Extract from the output of the Dominant Color extractor

The parameter Size represents the number of dominant colors found in the frame. The extraction of the "Spatial Coherency" can be consulted in [24]. The following pairs [Percentage, Values] stand for the percentage of dominance and the RGB combination values that this dominant color represents. Weka needs every file with the same number of inputs, so because the size can be variable, we have declared three new parameters DC_R, DC_G, DC_B as a single dominant color, where all the values are summed up as a weighted average of all those that have been found in each frame. The results can be found in Appendix VIII.A..

If we looked in detail to the results that we have obtained, the conclusion would not be that easy. We find three categories, for which the classifier improves notoriously (categories 1, 3 and 7), others which remain practically the same (categories 4, 5 and 6) and the remaining categories for which it does not work at all (categories 2 and 8). For these last ones, the main reason might be that there is no specific dominant color in the selected videos. Anyway, the general balance of the addition of this descriptor is mainly positive.

b) *Color Layout:* This MPEG-7 descriptor measures the similarity between different images based on the Color Layout. This procedure consists of five phases. The DCT + Quantification phase is covers the color space changing (from RGB to Y,Cb,Cr) and the resizing (from WxH to 8x8 for each plane). The result is a vector formed by 64 matrix coefficients and 6 and 5 bits for the DC and AC coefficients respectively.
The results can be found in Appendix VIII.B..

In this case, the results are much more stable than with the Dominant Color descriptor. The fact that the percentage of correctly detected instances remains always between 10% and 30% approximately in all the categories is an advantage because it means that the descriptors works fine with the classification of all the categories. It must be said though, that this results are not optimal and could of course still be improved. The categories that have obtained better results have been the first three ones (categories 1, 2 and 3).

4.6. Results comparative and evaluation

The best way to evaluate a system is by comparing the results obtained in each of the phases. In Table 5 it have been summarized the averaged percentages of correctly detected instances for each category. The evaluation concerns three stages: The individual descriptor classification (descriptors 1 to 6), the classification in the selected order (all descriptors in order: {2 → 3 → 6 → 4 → 5 → 1}) and the individual classification for the two new descriptors added (Dominant Color and color Layout).

Descriptors	Category								Average%
	1	2	3	4	5	6	7	8	
1	7.08	0.00	14.93	0.00	0.00	0.00	95.62	0.00	14.70
2	14.13	19.10	17.93	3.50	70.03	43.20	17.50	35.00	27.55
3	34.30	11.63	30.53	18.60	8.52	16.62	21.07	18.18	19.93
4	21.88	5.32	53.15	0.90	11.07	11.17	5.12	45.77	19.30
5	33.88	0.00	9.42	10.65	7.67	13.47	20.92	1.32	12.16
6	46.25	12.28	44.42	19.35	9.90	2.37	0.62	18.98	19.27
All descriptors	39.32	49.02	45.88	22.95	17.95	47.42	37.63	28.92	**36.14**
Dominant Color	46.30	0.00	43.72	29.45	39.30	13.88	37.50	1.88	26.50
Color Layout	19.95	16.72	17.77	17.98	16.72	13.85	21.53	18.27	17.85

Table 5 - Comparative table of results

As we can see, the results offer some peculiarities. Although the descriptor 1 (Face Detection) has the highest precision, 95.6% , when classifying videos from the category 7 (formula one) it works really bad classifying the rest of the categories. By the way, this is not so relevant

because this descriptor is the last to be added in the selected order so it does affect too much the final results. The average percentage of the six individual descriptors is 18.82%, is almost half the rate obtained if we accumulate all the descriptors in the selected order. If we compare this result with the ones obtained for the new descriptors, we can see that Dominant Color seems to be the descriptor that best suits the classification of sport videos. Color Layout is close to the average of the other descriptors, so the improvement is not much.

The conclusion obtained from the addition of these two descriptors denotes, in general terms, a little improvement compared to the last results we had. This means that sports categories really have a strong relationship with color characteristics. This can also be useful when classifying cartoons (where colors are usually very bright) or weather forecasts (where colors are usually white, blue and green), for instance.

Our system has been evaluated in both objective and subjective levels. From the objective point of view, the correct identification of the different genres of each of the videos has been measured with the software platform for automatic learning Weka. The obtained results are described in this and previous sections. The subjective evaluation has concerned the addition or elimination of determinate features according to the results obtained in the classification.

These results may seem low if we compare them to the ones obtained in other systems studied in the literature, but there are some important facts of our classification that must be taken into consideration. While their success rate is usually around 90% ours is much lower, around 36%. This is due to numerous reasons. First, we must take into account the importance of the selected database used in each case. As discussed in Chapter 3, the choice of our dataset has attempted to be made as realistic as possible. This means that each video, though belonging to the same category, may have very different characteristics (eg, videos of Roland Garós, depressed land, or Wimbledon, grass; both are tennis videos), which translates into more difficulty in finding a pattern of descriptors that characterize in a general way each class individually. Also be borne in mind that the videos used are mostly summaries of matches or competitions, which in turn contributes to a continuous heterogeneity of the structure of the videos.

Moreover, as already mentioned above, this work has focused on verifying the performance of a classifier from the descriptors point of view, nothing to do with the optimization of the SVM. To minimize programming costs, we have adapted an extractor of features destined to the summarization of on-line news bulletins to our purpose. Although these features prove to adapt well to sports genre, if the classifier had been developed from the outset is likely that some of these descriptors may had been replaced by more efficient ones, for example MPEG-7 Motion Activity.

Nevertheless, these results can still be improved. Some future lines of investigation could be the addition of new descriptors such as Motion Activity or the recognition of audio and text features. All this future work will be explained in detail in Chapter 5.

CHAPTER 5. CONCLUSIONS AND FUTURE WORK

5.1. Conclusions

Recently, the automatic classification of video has become a field of great interest due to the massive creation of multimedia content produced in tools such as the Internet. Channels like Youtube offer the user the possibility to become a constant producer of data, thus contributing to an overcrowding that has to be classified before it becomes uncontrollable. Video search by content or by subject is a problem we must face and which is currently under investigation to further improve the results obtained so far.

The development of systems like the one proposed in this work offers an unlimited number of benefits to efficiently handle with large amounts of data, as in the case of TV broadcasters, which would make profit of this kind of systems to facilitate the retrieval of video in an advanced and quick way, thus reducing costs of production and edition. Another gain for the entertainment industry is the application on services such as video-on-demand (VoD).

If we make a review of the results and we compare them with of the systems named in the literature, we could have made the mistake of thinking that the procedure used in this work has not been thoroughly debugged. Digging in those works we have concluded that their good results are due to the use of ad-hoc databases, that is, a set of videos carefully selected for one specific purpose and therefore best suited to the developed system. The problem in our case is that the database has been randomly chosen, so the videos do not necessarily have to follow any similar pattern in their characteristics, thus hampering the process of classification. The procedure followed in this work has advantages and disadvantages. Although the results are lower than in other works, they are more realistic, since when classifying videos, for example in the Internet, the variety of the content is infinite. The fact that we have used videos that are mainly summaries of matches and competitions also helps make it more difficult to generate a single vector of features for each category. To make a fair comparison, it should be noted that the content set used in these works has been really challenging to classify because of its realism. Some examples can be found at TrecVid. In this case, a success rate of 36% can be considered as a reasonably good result. On the other hand, these percentage can yet be improved.

5.2. Future work

First of all, and based on the results, as future work we should consider that this project is open to many improvements. To begin with, it is important to take into account that the implementation of a complete and better classified database can provide better classification results. The selection of categories that are easy to distinct, even manually, is not an easy job because the opinion of each user can be very different, so this should also be taken into consideration.

Automatic classification of video by genre is a broad field that allows a large number of lines of research. The possibility of experimenting with different learning algorithms, such as the K-Nearest Neighbors (K-NN) or the Parallel Neural Networks, can offer different views on the results and their possible improvements.

The main objective for the future is to develop a system robust enough to allow an extension to new genres and subgenres without losing efficiency on the classification. To this make this happen, it seems imperative the addition of new descriptors that are able to maintain a distinguished and structured category taxonomy. In this case, since in this work we have used only visual descriptors, we propose the development of a system that adds audio, speech recognition and text descriptors. For the classification of sport videos it can result really useful because there is a huge list of key words, both written (in the players' shirts, billboards or scoreboards) or spoken (by speakers or participants), which would be crucial in deciding one category or another . On the other hand, we must also consider the potential drawbacks of these descriptors, such as the poor quality of the audio or the image on a video, as well as the diversity of languages, which can make the development of the algorithm too complex. This can always be compensated with an appropriate database. In the literature, it is also recommended the investigation of "fuzzy classifiers" to fight limitations such as boundaries between categories that are not sharp enough.

One of the most useful end-user applications may be the creation of a semantic index of videos. Thus, it offers the possibility to analyze and annotate vast amounts of multimedia content in a standardized manner so the user can find quickly and efficiently only videos that are of his interest. Probably one of the biggest engines of the Internet on multimedia content is Youtube, which daily generates large amounts of video, so the implementation of an automatic video classification could truly ease the filtering of categories. This classification could be done in two ways: the first one would be to suggest the user a list of categories that the system had previously calculated from the values of the extracted descriptors. The second would be an automatic classification without consulting the user. The main difference between the two is the implicit human factor as the threshold between categories is not finely defined could become misleading given the subjectivity of each one.

It may also be useful for companies selling digital video recorders. In this case, the idea is to focus on the detection of commercials within a recording for later remove them and get the video stream immediately without commercial breaks. From the public institutions point of view, you can also use it to detect irregularities in implementing the laws like the number of minutes of commercials and unfair competition [10].

Last, but not least, one of the most useful applications of these systems of classification can be the detection of unsuitable content for minors (violence, pornography, etc.) on the Internet and its subsequent elimination.

REFERENCES

[1] V. Valdés, J.M. Martínez, **"On-line Video Abstract Generation of Multimedia News".** Multimedia Tools and Applications, Springer, ISSN 1380-7501 (Print) 1573-7721 (Online, March 2011) *(Digital Object Identifier*: 10.1007/s11042-011-0774-5)

[2] S. Fischer, R. Lienhart, W. Effelsberg, **"Automatic recognition of film genres"**, in Proc. of ACM Multimedia 1995

[3] Dimitrova N, Agnihotri L, Wei G **"Video classification based on HMM using text and Faces".** In Proc. of the European Conference on Signal Processing 2000

[4] Taskiran CM, Pollak I, Bouman CA, Delp EJ **"Stochastic models of video structure for program genre detection", in Proc**. of VLBV 2003

[5] Liu Z, Huang J, Wang Y (1998) **"Classification of TV programs based on audio information using hidden Markov model".** In Proc. of MMSP '98

[6] M.J. Roach, J.S.D. Mason, "**Video Genre Classification using Dynamics"**, in Proc. of XXXX 2001.

[7] J.C. San Miguel, **"Transmisión de secuencias de vídeo a tasa muy baja binaria adaptable basada en transmisión de descripción y síntesis"**, Juan Carlos San Miguel Avedillo, Proyecto Fin de Carrera, Titulación Ingeniero de Telecomunicación, Univ. Autónoma de Madrid, Escuela Politécnica Superior, Septiembre 2006

[8] L.Q. Xu, Y. Li , **"Video classification using spatial-temporal features and PCA", in Proc**. of ICME'03

[9] X. Yuan, W. Lai, T. Mei, X.S. Hua, X.Q. Wu, S. Li, **"Automatic video genre categorization using hierarchical SVM"**, in Proc. of ICIP 2006

[10] M. Montagnuolo, A. Messina, **"Parallel Neural Networks for Multimodal Video Genre Classification"**, Multimedia Tools and Applications, 41(1):125-159, 2009.

[11] B. Ionescu, C. Rasche, C. Vertan and P. Lambert, **"A Contour-Color-Action Approach to Automatic Classification of Several Common Video Genres".** In Proc. of AMR 2010.

[12] T.O. Ayodele, **"Types of Machine Learning Algorithms".** New Advances in Machine Learning, Yagang Zhang (Ed.) ISBN: 978-953-307-034-6. 2010.

[13] **"Tipos de clasificadores"**, http://es.wikipedia.org/wiki/Clasificadores_(matem%C3%A1tico)

[14] **"Rocchio classification",** http://en.wikipedia.org/wiki/Rocchio_Classification

[15] A. Albiol, M. Fullá, A. Albiol, L. Torres, **"Commercials detection using HMMs".** In Proc. of WIAMIS 2004

[16] R. Glasberg, A. Samour, K. Elazouzi, T. Sikora, **"Cartoon-recognition using video & audiodescriptors", in Proc**. of EUSIPCO2005

[17] T.I. Ianeva, A.P. de Vries, H. Rohrig , **"Detecting cartoons: a case study in automatic videogenre classification", in Proc**. of ICME'03

[18] J.M. Sánchez, X. Binefa, J. Vitriá, P. Radeva, **"Local color analysis for scene break detection applied to TV commercials recognition", in Proc**. of VISUAL '99

[19] S. Takagi, S. Hattori, K. Yokoyama, A. Kodate, H. Tominaga, **"Sports video categorizing method using camera motion parameters", in Proc**. of ICME'03

[20] C. Chang, C. Lin, **"LIBSVM : a library for support vector machines"**, ACM Transactions on Intelligent Systems and Technology, 2(27):1-, 2011. (Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm)

[21] G.Ciocca and R.Schettini, **"Dynamic key-frame extraction for video summarization"**, in Proc. of SPIE, vol. 5670, [On-line]. Available: http://link.aip.org/link/?PSI/5670/137/1

[22] R. Lienhart and J. Maydt, **"An extended set of haar-like features for rapid object detection",** In Proc. of ICIP 2002.

[23] M.J. Roach, **"Video genre classification".** PhD thesis, University of Wales Swansea, 2002

[24] J.C. San Miguel, **"Transmisión de secuencias de vídeo a tasa muy baja binaria adaptable basada en transmisión de descripción y síntesis"**, Juan Carlos San Miguel Avedillo, Proyecto Fin de Carrera, Titulación Ingeniero de Telecomunicación, Univ. Autónoma de Madrid, Escuela Politécnica Superior, Septiembre 2006

[25] D. Meyer, F. Leisch, and K. Hornik, **"The support vector machine under test"**, Neurocomputing, 55 (1-2):169-186, 2003.

[26] D. Brezeale, D.J. Cook, **"Automatic Video Classification: A Survey of the Literature"**, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 38(3), pp. 416-430, 2008.

[27] P.Q. Dinh, C. Dorai, S. Venkatesh**"Video genre categorization using audio wavelet coefficients", in Proc**. of ACCV2002

APPENDIX A.

CAPÍTULO 1. INTRODUCCIÓN

1.1. Introducción a la clasificación automática de vídeo

Los avances tecnológicos de esta era han llevado a la sociedad a hacer de la digitalización de la información una necesidad. El hecho de que los dispositivos electrónicos cada vez sean más baratos y de que en su mayoría, ofrezcan la posibilidad de conexión a Internet y de almacenamiento de datos, han catapultado que la tendencia de transformar documentos, música, periódicos, etcétera del papel al mundo multimedia hacia un crecimiento exponencial. Esta modernización se resume en un aumento considerable de la cantidad de datos en formato digital al que un usuario puede tener acceso, y por tanto, a una mayor dificultad a la hora de abstraer la información deseada.

La catalogación de todo este contenido es una tarea tediosa en la que habría que emplear no sólo un gran esfuerzo humano, sino también un gasto de tiempo considerable. Es por esto que se vuelve imprescindible tener un método que clasifique automáticamente y ordene todo el contenido almacenado. De esta manera, se hace más rápido y sencillo para el usuario encontrar únicamente los datos que son de su interés de acuerdo a un género en concreto.

La variedad de materiales audiovisuales es tan amplia que se considera que una clasificación por tipo de género puede ser una manera eficaz de ordenar dichos contenidos. Para bases de datos donde se almacenan películas, series o vídeos de cualquier clase, la realización de consultas como, por ejemplo, videos con contenidos parecidos, la utilización de este tipo de técnicas puede resultar especialmente útil. Para ello, se procura una extracción de un número determinado de características que no sea excesivamente grande y que a la vez sean suficientemente representativas para distinguir un único género.

En este proyecto se propone la clasificación automática por género de vídeos de deportes, en su mayoría resúmenes. En particular, para validar esta arquitectura de sistema, se ha desarrollado un marco experimental capaz de discernir entre vídeos de *fútbol, baloncesto, tenis, volley playa, snowboard, deportes acuáticos, fórmula 1* y *ciclismo.*

1.2. Motivación y objetivos

La inserción de clasificadores de vídeo automática en los sistemas reduce, incluso, elimina la intervención humana en dicho proceso, lo que puede ser muy útil para acelerar el procesamiento y la ordenación de enormes cantidades de datos.

El objetivo de este trabajo es desarrollar un algoritmo que sea capaz de clasificar automáticamente un número determinado de vídeos, especificando el género al que pertenecen. La experiencia del VPULab en competiciones como TRECVID y los avances en la generación de resúmenes de vídeos [1] han propiciado la motivación necesaria para seguir adelante con este proyecto. Este trabajo se basa en la "generación de resúmenes de vídeo en línea abstracta aplicado a noticias multimedia" de la que han reutilizado la mayoría de los métodos de extracción de descriptores.

El objetivo de este trabajo es implementar y mejorar el algoritmo citado anteriormente desarrollado para generar resúmenes de noticias multimedia a partir de videos on-line a otro método capaz de clasificar ocho tipos diferentes de videos deportivos. El estudio de la viabilidad del sistema, analizando sus características de acuerdo a la tecnología disponible en este momento, ha sido siempre el principal problema durante todo el análisis. Es importante tener en cuenta que cada video utilizado para este propósito ha sido resumido en una fase de pre-procesamiento. De esta manera, obtenemos lo que llamamos "Fotogramas clave", que son escenas muestreadas para que representen cada 25 segundos de vídeo, de las que los descriptores serán extraídos.

En resumen, para llevar a cabo este proyecto, nuestra metodología se ha centrado en:

- Selección de una base de datos robusta, para que ambos, el entrenamiento y las pruebas, nos procuren unos resultados válidos. En nuestro caso, los videos han sido extraídos de Youtube.

- Adquirir experiencia suficiente mediante numerosas simulaciones para poder desarrollar una solución eficaz al problema de la clasificación automática. Hemos basado nuestros esfuerzos en la mejora del trabajo previo.

- Desarrollar una fase de evaluación lo suficientemente detallada como para obtener conclusiones fiables sobre la viabilidad, los inconvenientes y la eficiencia de nuestro sistema.

Las mejoras en esta área debe extender la aplicación de estas técnicas no sólo a un alto nivel, es decir, a géneros completamente diferentes (noticias, anuncios, música, deportes…), sino también para distinguir clases dentro de dichos géneros, por ejemplo, entre películas de acción y comedias.

1.3. Descripción del documento

Este trabajo está estructurado de la siguiente manera: siguiendo a esta introducción, en el capítulo 2 se presenta el estado del arte de la clasificación de videos de deportes. En este apartado se incluyen, no sólo una visión general de las distintas técnicas estudiadas hasta el momento, sino también una explicación detallada de por qué hemos elegido para nuestro sistema la utilización de SVM. El capítulo 3 describe las decisiones que se han ido tomando para optimizar el sistema siguiendo un flujo específico de etapas de diseño. Por otra parte, en esta sección también se incluye una descripción de cada una de las características que van a ser utilizadas y sus razones. El Capítulo 4 presenta la parte de evaluación, incluyendo todos los experimentos que se han realizado durante todo el desarrollo. Los resultados que se han obtenido en cada una de las fases están incluidos en los apéndices. Finalmente, el capítulo 5 consagra el trabajo futuro en la investigación de lo previsto y las conclusiones sobre los resultados finales y la eficacia del sistema.

CAPÍTULO 5. CONCLUSIONES Y TRABAJO FUTURO

5.1. Conclusiones

Recientemente, la clasificación automática de video se ha convertido en un campo de gran interés debido a la creación masiva de contenidos multimedia en herramientas tan utilizadas como Internet. Canales como Youtube ofrecen la posibilidad al usuario de convertirse en un generador constante de documentos, contribuyendo así a una masificación que ha de ser clasificada para que no se descontrole. La búsqueda de vídeos por contenido o por temática es un problema que debemos afrontar y que actualmente está en proceso de investigación para seguir mejorando los resultados obtenidos hasta el momento.

El desarrollo de sistemas como el que se plantea en este proyecto suponen un número ilimitado de beneficios para el manejo eficiente de cantidades grandes de datos, como en el caso de las emisoras de televisión, que estaría orientado a facilitar el proceso de recuperación avanzada y rápida de vídeo, reduciendo así los costes de producción y edición. Otro de los beneficios para la industria del entretenimiento es la aplicación en servicios como el video-on-demand (VoD) o televisión a la carta.

Si se hace una revisión de los resultados y se comparan con los obtenidos en algunos de los sistemas mencionados en la literatura, se podría cometer el error de pensar que el procedimiento utilizado en este trabajo no ha sido totalmente depurado. Estudiando a fondo dichos resultados se descubre que sus buenos resultados se deben a la utilización de bases de datos ad-hoc, es decir, un conjunto de videos cuidadosamente seleccionados para un propósito específico y que, por lo tanto, se ajustan a la perfección al sistema desarrollado. El problema en nuestro caso es que la base de datos ha sido elegida al azar, por lo que los videos no siguen necesariamente ningún patrón común en sus características, lo que dificulta el proceso de clasificación. El procedimiento seguido en este trabajo tiene sus ventajas y sus desventajas. Aunque los resultados son más bajos que en otros trabajos, se puede decir que es debido a que se pretende tener un sistema lo más realista posible, puesto que a la hora de clasificar videos, por ejemplo, en Internet, la variedad de contenidos puede ser infinita. El hecho de que los videos utilizados sean principalmente resúmenes de partidos o competiciones contribuye a que sea más difícil generar un único vector de características para cada categoría. Para poder hacer una comparación justa, se deben buscar trabajos en los que la elección de contenidos haya sido un reto para la clasificación debido a su realismo. Algunos ejemplos se pueden encontrar en TrecVid. En este caso, una tasa de éxito del 36% puede ser considerado como un resultado razonablemente bueno. Por otro lado, estos porcentajes aún pueden mejorarse, como se comenta en el siguiente apartado.

5.2. Trabajo Futuro

En primer lugar, y basándonos en los resultados obtenidos, como trabajo futuro se debe considerar que este proyecto está abierto a numerosas mejoras. Para empezar, la implementación de una base de datos más completa y mejor clasificada puede ofrecer unos resultados mucho mejores. Así, también es importante determinar unas categorías que estén bien diferenciadas, de tal manera que a la hora de clasificar, incluso manualmente, no estén sujetas a dudas según la opinión del usuario.

La clasificación automática de vídeo por género es un campo muy amplio que permite un gran número de líneas de investigación. La posibilidad de probar con diferentes algoritmos de aprendizaje, como por ejemplo el de los K vecinos más cercanos (K-NN, K-Nearest Neighbors) o las redes paralelas neuronales, puede ofrecer diversos puntos de vista sobre los resultados obtenidos.

El principal objetivo para el futuro es desarrollar un sistema bien preparado que se pueda extender a nuevos géneros y subgéneros, teniendo en cuenta siempre que se mantenga la eficiencia de la clasificación. Para ello, se supone la adición de nuevos descriptores que sean capaces de mantener una taxonomía de categorías bien diferenciada. En este caso, puesto que en este trabajo sólo se han utilizado descriptores visuales, se propone el desarrollo de un sistema que añada descriptores de audio, reconocimiento del habla y texto. Para la clasificación de vídeos de deportes puede resultar muy útil porque hay palabras clave, tanto escritas en las camisetas de los jugadores, vallas publicitarias o marcadores, así como habladas por locutores o participantes, que serían determinantes a la hora de decidir una categoría u otra. Por otro lado, hay que tener en cuenta también los posibles inconvenientes de estos descriptores, como por ejemplo, la mala calidad del audio o la imagen de un vídeo o la diversidad de idiomas, que pueden hacer el desarrollo del algoritmo demasiado complicado si no se elige una base de datos adecuada. En trabajos anteriores, se recomienda la investigación de los clasificadores confusos ("fuzzy classifiers") para combatir limitaciones como unas fronteras que no sean lo suficientemente nítidas entre categorías.

Una de las aplicaciones más importantes puede ser la creación de un índice semántico de vídeos. De esta manera, se ofrece la posibilidad de analizar y anotar cantidades ingentes de contenido multimedia de una manera estandarizada pudiendo el usuario encontrar de forma rápida y eficiente únicamente los vídeos que sean de su interés. Uno de los motores más importantes de internet por la gran cantidad de contenidos multimedia que genera diariamente es Youtube, por lo que la implantación de un sistema de clasificación automática de vídeo podría facilitar mucho el filtrado de categorías. Dicha clasificación se podría hacer de dos maneras: la primera, sería sugerir al usuario una lista de categorías que el sistema detectaría según los valores de los descriptores extraídos y la segunda, sería una clasificación automática sin la consulta al usuario. La diferencia principal entre ambas sería el factor humano implícito, que puesto que el umbral entre categorías no está finamente definido podría dar lugar a error teniendo en cuenta la subjetividad de cada uno.

También puede resultar de gran utilidad para empresas de venta de aparatos de grabación de vídeo digitales (ej. TDT grabador). En este caso, la idea es centrarse en la detección de anuncios dentro de una grabación para posteriormente eliminarlos y obtener la secuencia de vídeo seguida sin pausas para publicidad. Desde el punto de vista del gobierno, se puede hacer uso también para detectar irregularidades en el cumplimiento de las leyes sobre número de minutos de publicidad o competencia desleal [10].

Por último, pero no por ello menos destacable, una de las aplicaciones más útiles de este tipo de sistemas de clasificación puede ser la detección de contenido no apto para menores (violencia, pornografía, etc.) en Internet y su consecuente eliminación.

APPENDIX B. Results

APPENDIX I

I. Binary SVM for the selection of an efficient Split Percentage

Table 6 - Binary SVM for the selection of an efficient Split Percentage

Class	Split_%	%_Correct	%_Incorrect	Precision		Recall	
				CATEGORY 1	CATEGORY 0	CATEGORY 1	CATEGORY 0
1	***20***	69.4	30.6	0.217	0.931	0.607	0.706
1	***40***	66.7	33.3	0.256	0.934	0.716	0.659
1	***60***	66.3	33.7	0.172	0.93	0.57	0.674
1	***80***	59.9	40.1	0.337	0.818	0.608	0.595
2	***20***	69.4	30.6	0.215	0.974	0.83	0.68
2	***40***	70.8	29.2	0.254	0.969	0.826	0.693
2	***60***	73.3	26.7	0.359	0.981	0.925	0.697
2	***80***	64.12	35.88	0.244	0.984	0.928	0.601
3	***20***	64.46	35.54	0.412	0.778	0.516	0.698
3	***40***	72.01	27.99	0.514	0.84	0.65	0.794
3	***60***	75.1	24.9	0.522	0.894	0.755	0.794
3	***80***	79.2	20.8	0	1	0	0.792
4	***20***	68.98	31.02	0.063	0.968	0.466	0.7
4	***40***	60.7	39.3	0.066	0.965	0.559	0.61
4	***60***	70.9	29.1	0.039	0.984	0.5	0.714
4	***80***	59.8	40.2	0	1	0	0.598
5	***20***	69.6	30.4	0.344	0.92	0.734	0.687
5	***40***	74.5	25.5	0.371	0.949	0.8	0.734
5	***60***	79.2	20.8	0.465	0.919	0.688	0.816
5	***80***	70.4	29.6	0.549	0.823	0.707	0.702
6	***20***	66.1	33.9	0.123	0.914	0.403	0.689
6	***40***	72.7	27.3	0.179	0.951	0.597	0.739
6	***60***	70.3	29.7	0.133	0.962	0.613	0.71
6	***80***	73.8	26.2	0.077	0.96	0.388	0.757
7	***20***	69.3	30.7	0.096	0.981	0.713	0.692
7	***40***	72.8	27.2	0.121	0.978	0.693	0.73
7	***60***	74.6	25.4	0.189	0.965	0.678	0.752
7	***80***	57.1	42.9	0.178	0.957	0.804	0.542
8	***20***	61.7	38.3	0.18	0.904	0.553	0.672
8	***40***	63.1	36.9	0.147	0.916	0.508	0.646
8	***60***	68.1	31.9	0.231	0.952	0.738	0.679
8	***80***	69.8	30.2	0.273	0.967	0.839	0.677

APPENDIX II

II. a) Binary SVM. Analysis of the results using the "False Label" method.

Table 7 - Binary SVM. Analysis of the results using the "False Label" method.

Category	Clase Training	Correct Instances %	Incorrect Instances %
1	1	34.9	65.1
	2	18.4	81.6
	3	36.2	63.8
	4	1.2	98.8
	5	21.5	78.5
	6	51.5	48.5
	7	35.6	64.4
	8	35.5	64.5

Category	Clase Training	Correct Instances %	Incorrect Instances %
2	1	49.1	50.9
	2	42.5	57.5
	3	23.2	76.8
	4	11.5	88.5
	5	26.9	73.1
	6	34.7	65.3
	7	31.8	68.2
	8	45.5	54.5

Category	Clase Training	Correct Instances %	Incorrect Instances %
3	1	34.9	65.1
	2	30.1	69.9
	3	19.4	80.6
	4	13.7	86.3
	5	22.8	77.2
	6	43.1	56.9
	7	28.6	71.4
	8	30.9	69.1

Category	Clase Training	Correct Instances %	Incorrect Instances %
4	1	26.6	73.4
	2	25.8	74.2
	3	28.1	71.9
	4	31.9	68.1
	5	35.4	64.6
	6	38.2	61.8
	7	36.4	63.6
	8	26.1	73.9

Category	Clase Training	Correct Instances %	Incorrect Instances %
5	1	27.5	72.5
	2	24.3	75.7
	3	32.1	67.9
	4	33.7	66.3
	5	49.3	50.7
	6	27.4	72.6
	7	45.2	54.8
	8	19.4	80.6

Category	Clase Training	Correct Instances %	Incorrect Instances %
6	1	21.7	78.3
	2	20.5	79.5
	3	37.6	62.4
	4	43.1	56.9
	5	47.1	52.9
	6	40.8	59.2
	7	38.5	61.5
	8	15.2	84.8

Category	Clase Training	Correct Instances %	Incorrect Instances %
7	1	35.2	64.8
	2	24.9	75.1
	3	30.1	69.9
	4	38.7	61.3
	5	42.6	57.4
	6	33.7	66.3
	7	50.4	49.6
	8	31.9	68.1

Category	Clase Training	Correct Instances %	Incorrect Instances %
8	1	34.2	65.8
	2	23	77
	3	25.9	74.1
	4	35.9	64.1
	5	39.1	60.9
	6	35.9	64.1
	7	50.5	49.5
	8	34.6	65.4

II. b) Multiclass SVM. Analysis of the results using the "False Label" method to select an efficient Split Percentage.

Table 8 - Multiclass SVM. Analysis of the results using the "False Label" method to select an efficient Split Percentage.

20%	1	2	3	4	5	6	7	8
1	3	13	13	2	45	66	9	12
2	40	24	0	2	8	0	1	4
3	0	11	2	35	34	34	23	0
4	7	8	34	50	20	4	1	9
5	51	25	61	21	137	4	19	11
6	2	37	65	112	128	43	3	10
7	63	103	0	0	0	0	24	0
8	224	72	19	7	2	0	6	5

40%	1	2	3	4	5	6	7	8
1	17	9	53	21	7	13	39	4
2	53	11	1	4	6	0	0	4
3	3	3	9	15	3	83	23	0
4	1	7	23	71	10	13	2	6
5	68	15	76	45	94	1	21	9
6	14	34	104	118	14	98	11	7
7	4	2	0	0	0	1	280	23
8	11	4	59	23	6	95	91	46

60%	1	2	3	4	5	6	7	8
1	26	3	20	0	3	37	69	5
2	40	27	0	0	11	0	1	0
3	5	2	7	6	3	75	40	1
4	2	2	3	84	24	10	3	5
5	72	6	52	11	133	3	43	9
6	19	8	29	120	84	97	26	17
7	4	2	0	0	0	0	272	32
8	17	3	42	33	0	59	125	56

80%	1	2	3	4	5	6	7	8
1	37	6	58	0	11	7	40	4
2	34	33	0	0	11	0	1	0
3	5	3	25	2	6	72	24	2
4	3	1	4	49	59	10	3	4
5	60	12	50	10	155	7	27	8
6	15	18	38	75	90	147	10	7
7	1	4	0	0	0	0	104	201
8	10	2	54	0	13	23	20	213

50%	1	2	3	4	5	6	7	8
1	25	7	66	8	20	11	24	2
2	47	17	0	0	11	0	0	0
3	8	2	63	17	5	22	22	0
4	3	6	18	73	15	9	2	7
5	68	14	40	17	142	1	40	7
6	19	26	60	135	75	67	11	7
7	5	4	0	0	0	0	292	9
8	16	3	73	26	11	69	95	42

100%	1	2	3	4	5	6	7	8
1	29	6	60	1	12	3	49	3
2	41	26	0	0	12	0	0	0
3	5	1	29	9	7	58	30	0
4	2	3	5	68	39	8	3	5
5	65	10	41	12	163	6	24	8
6	15	19	28	84	89	140	18	7
7	2	2	0	0	0	0	175	131
8	14	3	53	9	8	37	55	56

APPENDIX III

III. Multiclass SVM. Analysis of the results using the "Median Window" filter

This procedure is very robust and has only been applied to three videos per category. It has been checked that the addition of more videos does not vary much the results.

Table 9 . Analysis of the results using the "Median Window" filter

Category	Video	Majority class		Confused with	
1	**7_1**	1	42,1%	-	-
	8_1	1	44%	-	-
	9_1	1	55,5%	-	-
2	**7_2**	2	13,6%	1	51,5%
	8_2	2	30,1%	1	38,5%
	9_2	2	54,2%	-	-
3	**7_3**	3	35,4%	-	-
	8_3	3	65,5%	-	-
	9_3	3	35,2%	1	52,9%
4	**7_4**	4,5,6	15,8%	1	36,8%
	8_4	4	22,2%	1	55,5%
	9_4	4,5	23,9%	1	30,4%
5	**7_5**	5	15,1%	7	41,7%
	8_5	2,5,7	18,1%	1	36,3%
	9_5	5	9,09%	4	33,3%
6	**7_6**	6	51,4%	-	-
	8_6	6	70%	-	-
	9_6	6	47,1%	-	-
7	**7_7**	7	21,1%	1,3	26,3%
	8_7	7	68,9%	-	-
	9_7	7	9,09%	1	40,1%
8	**7_8**	8	29,6%	1	44,4%
	8_8	8	29,6%	-	-
	9_8	8	16,7%		38,1%

The conclusion obtained from the results is that the most common confusion is made with category 1, football. This might be related to the looseness of the features that does not adapt that good to all the categories. The training videos chosen for the category 1 may be too general for this features and the machine learns just global patterns that can be found in any other category too. To solve this problem we have changed the videos of category one and we have added two more features. The improvement in the results can be seen in the following appendixes.

APPENDIX IV

IV. Multiclass SVM. Classification using the descriptors individually.

CATEGORY 1

VIDEO 7_1		
	Correctly detected %	Incorrectly detected %
1	11.34	88.66
2	13.4	86.6
3	19.58	80.42
4	13.4	86.6
5	15.4	84.6
6	27.8	72.2

VIDEO 11_1		
	Correctly detected %	Incorrectly detected %
1	1.3	98.7
2	8.4	91.6
3	30.4	69.6
4	9.7	90.3
5	27.2	72.8
6	36.6	63.4

VIDEO 10_1		
	Correctly detected %	Incorrectly detected %
1	15.12	84.88
2	9.3	90.7
3	39.53	60.47
4	37.2	62.8
5	46.5	53.5
6	51.2	48.8

VIDEO 9_1		
	Correctly detected %	Incorrectly detected %
1	3.4	96.6
2	21.2	78.8
3	31.9	68.1
4	34.4	65.6
5	39.1	60.9
6	57.2	42.8

VIDEO 8_1		
	Correctly detected %	Incorrectly detected %
1	4	96
2	5.6	94.4
3	34.4	65.6
4	11.2	88.8
5	24.8	75.2
6	38.4	61.6

VIDEO 12_1		
	Correctly detected %	Incorrectly detected %
1	7.3	92.7
2	26.9	73.1
3	50	50
4	25.4	74.6
5	50.3	49.7
6	66.3	33.7

CATEGORY 2

VIDEO 7_2		
	Correctly detected %	Incorrectly detected %
1	0	100
2	18.7	81.3
3	0	100
4	9.1	90.9
5	0	100
6	15.7	84.3

VIDEO 11_2		
	Correctly detected %	Incorrectly detected %
1	0	100
2	28.7	71.3
3	15.6	84.4
4	4.1	95.9
5	0	100
6	0	100

VIDEO 10_2		
	Correctly detected %	Incorrectly detected %
1	0	100
2	11.7	88.3
3	0	100
4	5.8	94.2
5	0	100
6	7.1	92.9

VIDEO 9_2		
	Correctly detected %	Incorrectly detected %
1	0	100
2	17.7	82.3
3	0	100
4	5.6	94.4
5	0	100
6	28.2	71.8

VIDEO 8_2		
	Correctly detected %	Incorrectly detected %
1	0	100
2	17.7	82.3
3	0	100
4	5.2	94.8
5	0	100
6	17.2	82.8

VIDEO 12_2		
	Correctly detected %	Incorrectly detected %
1	0	100
2	20.1	79.9
3	54.2	45.8
4	2.1	97.9
5	0	100
6	5.5	94.5

CATEGORY 3

VIDEO 7_3		
	Correctly detected %	Incorrectly detected %
1	21.5	78.5
2	28.9	71.1
3	0	100
4	25.3	74.7
5	0	100
6	26.1	73.9

VIDEO 11_3		
	Correctly detected %	Incorrectly detected %
1	4.6	95.4
2	8.1	91.9
3	47.1	52.9
4	55.9	44.1
5	25	75
6	57.5	42.5

VIDEO 10_3		
	Correctly detected %	Incorrectly detected %
1	11.9	88.1
2	6.1	93.9
3	43.3	56.7
4	53.9	46.1
5	15.7	84.3
6	49.9	50.1

VIDEO 9_3		
	Correctly detected %	Incorrectly detected %
1	7.6	92.4
2	9.4	90.6
3	44.5	55.5
4	57.5	42.5
5	8.6	91.4
6	54.7	45.3

VIDEO 8_3		
	Correctly detected %	Incorrectly detected %
1	22.2	77.8
2	52.2	47.8
3	0	100
4	57.9	42.1
5	0	100
6	12.9	87.1

VIDEO 12_3		
	Correctly detected %	Incorrectly detected %
1	21.8	78.2
2	2.9	97.1
3	48.3	51.7
4	68.4	31.6
5	7.2	92.8
6	65.4	34.6

CATEGORY 4

VIDEO 7_4		
	Correctly detected %	Incorrectly detected %
1	0	100
2	5.1	94.9
3	0	100
4	3.1	96.9
5	0	100
6	7.1	92.9

VIDEO 11_4		
	Correctly detected %	Incorrectly detected %
1	0	100
2	0	100
3	12.6	87.4
4	0	100
5	1.4	98.6
6	4	96

VIDEO 10_4		
	Correctly detected %	Incorrectly detected %
1	0	100
2	0	100
3	1	99
4	0	100
5	2.1	97.9
6	3	97

VIDEO 9_4		
	Correctly detected %	Incorrectly detected %
1	0	100
2	0	100
3	64.4	35.6
4	0	100
5	43.1	56.9
6	73.1	26.9

VIDEO 8_4		
	Correctly detected %	Incorrectly detected %
1	0	100
2	15.9	84.1
3	0	100
4	2.3	97.7
5	0	100
6	4.5	95.5

VIDEO 12_4		
	Correctly detected %	Incorrectly detected %
1	0	100
2	0	100
3	33.6	66.4
4	0	100
5	17.3	82.7
6	24.4	75.6

CATEGORY 5

VIDEO 7_5		
	Correctly detected %	Incorrectly detected %
1	0	100
2	77.7	22.3
3	12.7	87.3
4	3.3	96.7
5	9.8	90.2
6	10.1	89.9

VIDEO 11_5		
	Correctly detected %	Incorrectly detected %
1	0	100
2	92.3	7.7
3	5.6	94.4
4	0	100
5	0	100
6	0	100

VIDEO 10_5		
	Correctly detected %	Incorrectly detected %
1	0	100
2	74.3	25.7
3	10.4	89.6
4	7.1	92.9
5	9.5	90.5
6	12.1	87.9

VIDEO 9_5		
	Correctly detected %	Incorrectly detected %
1	0	100
2	36.8	63.2
3	4.1	95.9
4	15.8	84.2
5	3.5	96.5
6	9.9	90.1

VIDEO 8_5		
	Correctly detected %	Incorrectly detected %
1	0	100
2	82.4	17.6
3	2.6	97.4
4	18.2	81.8
5	0.5	99.5
6	5.3	94.7

VIDEO 12_5		
	Correctly detected %	Incorrectly detected %
1	0	100
2	56.7	43.3
3	15.7	84.3
4	22	78
5	22.7	77.3
6	22	78

CATEGORY 6

VIDEO 7_6		
	Correctly detected %	Incorrectly detected %
1	0	100
2	34.9	65.1
3	20.2	79.8
4	5.5	94.5
5	8.7	91.3
6	1.1	98.9

VIDEO 11_6		
	Correctly detected %	Incorrectly detected %
1	0	100
2	36.3	63.7
3	16.1	83.9
4	12.3	87.7
5	6.3	93.7
6	2.4	97.6

VIDEO 10_6		
	Correctly detected %	Incorrectly detected %
1	0	100
2	42.9	57.1
3	19.1	80.9
4	11.6	88.4
5	12.1	87.9
6	5.1	94.9

VIDEO 9_6		
	Correctly detected %	Incorrectly detected %
1	0	100
2	24.7	75.3
3	4.7	95.3
4	10.6	89.4
5	36.5	63.5
6	3.5	96.5

VIDEO 8_6		
	Correctly detected %	Incorrectly detected %
1	0	100
2	48.3	51.7
3	18.5	81.5
4	13.9	86.1
5	5.3	94.7
6	0	100

VIDEO 12_6		
	Correctly detected %	Incorrectly detected %
1	0	100
2	72.1	27.9
3	21.1	78.9
4	13.1	86.9
5	11.9	88.1
6	2.1	97.9

CATEGORY 7

VIDEO 7_7		
	Correctly detected %	Incorrectly detected %
1	96.1	3.9
2	7.8	92.2
3	18.4	81.6
4	0	100
5	1.9	98.1
6	0	100

VIDEO 11_7		
	Correctly detected %	Incorrectly detected %
1	97.4	2.6
2	18.3	81.7
3	34.9	65.1
4	3.5	96.5
5	37.1	62.9
6	0	100

VIDEO 10_7		
	Correctly detected %	Incorrectly detected %
1	88.1	11.9
2	7.1	92.9
3	26.2	73.8
4	0	100
5	7.1	92.9
6	0	100

VIDEO 9_7		
	Correctly detected %	Incorrectly detected %
1	93.2	6.8
2	35.4	64.6
3	17.7	82.3
4	12.2	87.8
5	19.1	80.9
6	2.1	97.9

VIDEO 8_7		
	Correctly detected %	Incorrectly detected %
1	98.9	1.1
2	27.5	72.5
3	21.8	78.2
4	13.5	86.5
5	56.9	43.1
6	1.6	98.4

VIDEO 12_7		
	Correctly detected %	Incorrectly detected %
1	100	0
2	8.9	91.1
3	7.4	92.6
4	1.5	98.5
5	3.4	96.6
6	0	100

CATEGORY 8

VIDEO 7_8		
	Correctly detected %	Incorrectly detected %
1	0	100
2	39.1	60.9
3	19.9	80.1
4	48.2	51.8
5	0	100
6	23.9	76.1

VIDEO 11_8		
	Correctly detected %	Incorrectly detected %
1	0	100
2	40.1	59.9
3	36.2	63.8
4	72.8	27.2
5	0	100
6	17.4	82.6

VIDEO 10_8		
	Correctly detected %	Incorrectly detected %
1	0	100
2	53.6	46.4
3	9.3	90.7
4	36.1	63.9
5	7.9	92.1
6	18.1	81.9

VIDEO 9_8		
	Correctly detected %	Incorrectly detected %
1	0	100
2	20.3	79.7
3	19.4	80.6
4	57.1	42.9
5	0	100
6	17.5	82.5

VIDEO 8_8		
	Correctly detected %	Incorrectly detected %
1	0	100
2	38.5	61.5
3	12.2	87.8
4	33.5	66.5
5	0	100
6	16.7	83.3

VIDEO 12_8		
	Correctly detected %	Incorrectly detected %
1	0	100
2	18.4	81.6
3	12.1	87.9
4	26.9	73.1
5	0	100
6	20.3	79.7

Table 10 - Multiclass SVM. Classification using the descriptors individually.

APPENDIX V

V. Selection of an efficient order of the descriptors based on the results from the individual descriptors classification.

Table 11 - Selection of an efficient order of the descriptors based on the results from the individual descriptors classification.

CATEGORY 1						
	7_1	8_1	9_1	10_1	11_1	12_1
1	6	6	6	6	6	6
2	3	3	5	5	3	5
3	5	5	4	3	5	3
4	4,2	4	3	4	4	2
5	1	2	2	1	2	4
6		1	1	2	1	1

CATEGORY 2						
	7_1	8_1	9_1	10_1	11_1	12_1
1	2	2	6	2	2	3
2	6	6	2	6	3	2
3	4	4	4	4	4	6
4	1,3,5	1,3,5	1,3,5	1,3,5	1,6,5	4
5						1,5
6						

CATEGORY 3						
	7_1	8_1	9_1	10_1	11_1	12_1
1	2	4	4	4	6	4
2	6	2	6	6	4	6
3	4	1	3	3	3	3
4	1	6	2	5	5	1
5	3,5	3,5	5	1	2	5
6			1	2	1	2

CATEGORY 4						
	7_1	8_1	9_1	10_1	11_1	12_1
1	6	2	6	6	3	3
2	2	6	3	5	6	6
3	4	4	5	3	5	5
4	1,3,5	1,3,5	1,2,4	1,2,4	1,2,4	1,2,4
5						
6						

CATEGORY 5						
	7_1	8_1	9_1	10_1	11_1	12_1
1	2	2	2	2	2	2
2	3	4	4	6	3	5
3	6	6	6	3	1,4,5,6	4,6
4	5	3	3	5		3
5	4	1,5	5	4		1
6	1		1	1		

CATEGORY 6						
	7_1	8_1	9_1	10_1	11_1	12_1
1	2	2	5	2	2	2
2	3	3	2	3	3	3
3	5	4	4	5	4	4
4	4	5	3	4	5	5
5	6	1,6	6	6	6	6
6	1		1	1	1	1

CATEGORY 7						
	7_1	8_1	9_1	10_1	11_1	12_1
1	1	1	1	1	1	1
2	3	5	2	3	5	2
3	2	2	5	2,5	3	3
4	5	3	3	4,6	2	5
5	4,6	4	4		4	4
6		6	6		6	6

CATEGORY 8						
	7_1	8_1	9_1	10_1	11_1	12_1
1	4	2	4	2	4	4
2	2	4	2	4	2	6
3	6	6	3	6	3	2
4	3	3	6	3	6	3
5	1,5	1,5	1,5	5	1,5	1,5
6				1		

Cells in pink represent the descriptors that have obtained a 0% of corrected instances. This means that they are not relevant at all for that videos.

APPENDIX VI

VI. Multiclass SVM. Accumulative application of descriptors in the established order.

CATEGORY 1

VIDEO 7_1		
	Correctly detected %	Incorrectly detected %
2	13.4	86.6
2_3	17.5	82.5
2_3_6	29.8	70.2
2_3_6_4	28.9	71.1
2_3_6_4_5	30.9	69.1
2_3_6_4_5_1	29.8	70.2

VIDEO 11_1		
	Correctly detected %	Incorrectly detected %
2	8.4	91.6
2_3	40.5	59.5
2_3_6	45.6	54.4
2_3_6_4	34.3	65.7
2_3_6_4_5	38.3	61.7
2_3_6_4_5_1	29.5	70.5

VIDEO 10_1		
	Correctly detected %	Incorrectly detected %
2	9.3	90.7
2_3	6.9	93.1
2_3_6	43.1	56.9
2_3_6_4	53.5	46.5
2_3_6_4_5	56.9	43.1
2_3_6_4_5_1	56.9	43.1

VIDEO 9_1		
	Correctly detected %	Incorrectly detected %
2	21.2	78.8
2_3	10.9	89.1
2_3_6	28.4	71.6
2_3_6_4	35.5	64.5
2_3_6_4_5	37.4	62.6
2_3_6_4_5_1	37.6	62.4

VIDEO 8_1		
	Correctly detected %	Incorrectly detected %
2	5.6	94.4
2_3	6.4	93.6
2_3_6	28.8	71.2
2_3_6_4	29.6	70.4
2_3_6_4_5	33.6	66.4
2_3_6_4_5_1	33.6	66.4

VIDEO 12_1		
	Correctly detected %	Incorrectly detected %
2	26.9	73.1
2_3	55.2	44.8
2_3_6	60.9	39.1
2_3_6_4	59.8	40.2
2_3_6_4_5	63.9	36.1
2_3_6_4_5_1	48.5	51.5

CATEGORY 2

VIDEO 7_2		
	Correctly detected %	Incorrectly detected %
2	18.7	81.3
2_3	4.5	95.5
2_3_6	37.1	62.9
2_3_6_4	24.4	75.6
2_3_6_4_5	32.5	67.5
2_3_6_4_5_1	27.1	72.9

VIDEO 11_2		
	Correctly detected %	Incorrectly detected %
2	28.7	71.3
2_3	56.3	43.7
2_3_6	56.9	43.1
2_3_6_4	64.7	35.3
2_3_6_4_5	65.3	34.7
2_3_6_4_5_1	67.1	32.9

VIDEO 10_2		
	Correctly detected %	Incorrectly detected %
2	11.7	88.3
2_3	3.5	96.5
2_3_6	43.5	56.5
2_3_6_4	45.8	54.2
2_3_6_4_5	54.1	45.9
2_3_6_4_5_1	53.2	46.8

VIDEO 9_2		
	Correctly detected %	Incorrectly detected %
2	17.7	82.3
2_3	10.4	89.6
2_3_6	49.2	50.8
2_3_6_4	50	50
2_3_6_4_5	54.1	45.9
2_3_6_4_5_1	53.2	46.8

VIDEO 8_2		
	Correctly detected %	Incorrectly detected %
2	17.7	82.3
2_3	6.7	93.3
2_3_6	55.1	44.9
2_3_6_4	34.9	65.1
2_3_6_4_5	45.7	54.3
2_3_6_4_5_1	35.9	64.1

VIDEO 12_2		
	Correctly detected %	Incorrectly detected %
2	20.1	79.9
2_3	54.2	45.8
2_3_6	49.3	50.7
2_3_6_4	55.5	44.5
2_3_6_4_5	61.8	38.2
2_3_6_4_5_1	57.6	42.4

CATEGORY 3

VIDEO 7_3		
	Correctly detected %	Incorrectly detected %
2	28.9	71.1
2_3	21.9	78.1
2_3_6	18.2	81.8
2_3_6_4	31.6	68.4
2_3_6_4_5	34.4	65.6
2_3_6_4_5_1	33.7	66.3

VIDEO 11_3		
	Correctly detected %	Incorrectly detected %
2	8.1	91.9
2_3	22.6	77.4
2_3_6	34.4	65.6
2_3_6_4	45.7	54.3
2_3_6_4_5	47.8	52.2
2_3_6_4_5_1	40.1	59.9

VIDEO 10_3		
	Correctly detected %	Incorrectly detected %
2	6.1	93.9
2_3	12.9	87.1
2_3_6	15.4	84.6
2_3_6_4	43.5	56.5
2_3_6_4_5	47.1	52.9
2_3_6_4_5_1	39.5	60.5

VIDEO 9_3		
	Correctly detected %	Incorrectly detected %
2	9.4	90.6
2_3	10.9	89.1
2_3_6	14.9	85.1
2_3_6_4	40.1	59.9
2_3_6_4_5	45.8	54.2
2_3_6_4_5_1	35.3	64.7

VIDEO 8_3		
	Correctly detected %	Incorrectly detected %
2	52.1	47.9
2_3	22.4	77.6
2_3_6	19.9	80.1
2_3_6_4	66.2	33.8
2_3_6_4_5	65.6	34.4
2_3_6_4_5_1	63.9	36.1

VIDEO 12_3		
	Correctly detected %	Incorrectly detected %
2	2.9	97.1
2_3	13.7	86.3
2_3_6	21.8	78.2
2_3_6_4	64.5	35.5
2_3_6_4_5	65.8	34.2
2_3_6_4_5_1	62.8	37.2

CATEGORY 4

VIDEO 7_4		
	Correctly detected %	Incorrectly detected %
2	5.1	94.9
2_3	2.04	97.96
2_3_6	12.2	87.8
2_3_6_4	12.2	87.8
2_3_6_4_5	14.3	85.7
2_3_6_4_5_1	18.4	81.6

VIDEO 11_4		
	Correctly detected %	Incorrectly detected %
2	0	100
2_3	7.9	92.1
2_3_6	5.4	94.6
2_3_6_4	3.1	96.9
2_3_6_4_5	2.7	97.3
2_3_6_4_5_1	2.5	97.5

VIDEO 10_4		
	Correctly detected %	Incorrectly detected %
2	0	100
2_3	0	100
2_3_6	2.8	97.2
2_3_6_4	2.3	97.7
2_3_6_4_5	2.1	97.9
2_3_6_4_5_1	2.3	97.7

VIDEO 9_4		
	Correctly detected %	Incorrectly detected %
2	0	100
2_3	55.1	44.9
2_3_6	66.6	33.4
2_3_6_4	59.9	40.1
2_3_6_4_5	61.8	38.2
2_3_6_4_5_1	62.2	37.8

VIDEO 8_4		
	Correctly detected %	Incorrectly detected %
2	15.9	84.1
2_3	2.3	97.7
2_3_6	40.9	59.1
2_3_6_4	38.6	61.4
2_3_6_4_5	40.9	59.1
2_3_6_4_5_1	20.5	79.5

VIDEO 12_4		
	Correctly detected %	Incorrectly detected %
2	0	100
2_3	20.8	79.2
2_3_6	32.5	67.5
2_3_6_4	25.1	74.9
2_3_6_4_5	26.5	73.5
2_3_6_4_5_1	31.8	68.2

CATEGORY 5

VIDEO 7_5		
	Correctly detected %	Incorrectly detected %
2	77.7	22.3
2_3	84.1	15.9
2_3_6	38.6	61.4
2_3_6_4	25.8	74.2
2_3_6_4_5	22.3	77.7
2_3_6_4_5_1	16.5	83.5

VIDEO 11_5		
	Correctly detected %	Incorrectly detected %
2	92.3	7.7
2_3	9.1	90.9
2_3_6	6.5	93.5
2_3_6_4	2.6	97.4
2_3_6_4_5	3.5	96.5
2_3_6_4_5_1	5.1	94.9

VIDEO 10_5		
	Correctly detected %	Incorrectly detected %
2	74.3	25.7
2_3	81.7	18.3
2_3_6	40.6	59.4
2_3_6_4	28.6	71.4
2_3_6_4_5	18.7	81.3
2_3_6_4_5_1	20.3	79.7

VIDEO 9_5		
	Correctly detected %	Incorrectly detected %
2	36.8	63.2
2_3	59.9	40.1
2_3_6	9.6	90.4
2_3_6_4	8.5	91.5
2_3_6_4_5	8.2	91.8
2_3_6_4_5_1	9.6	90.4

VIDEO 8_5		
	Correctly detected %	Incorrectly detected %
2	82.4	17.6
2_3	90.4	9.6
2_3_6	27.8	72.2
2_3_6_4	27.3	72.7
2_3_6_4_5	19.8	80.2
2_3_6_4_5_1	20.9	79.1

VIDEO 12_5		
	Correctly detected %	Incorrectly detected %
2	56.7	43.3
2_3	18	82
2_3_6	29	71
2_3_6_4	36	64
2_3_6_4_5	35.7	64.3
2_3_6_4_5_1	35.3	64.7

CATEGORY 6

VIDEO 7_6		
	Correctly detected %	Incorrectly detected %
2	34.9	65.1
2_3	44.3	55.7
2_3_6	52.5	47.5
2_3_6_4	48.1	51.9
2_3_6_4_5	48.1	51.9
2_3_6_4_5_1	49.7	50.3

VIDEO 11_6		
	Correctly detected %	Incorrectly detected %
2	36.3	63.7
2_3	37.8	62.2
2_3_6	32.1	67.9
2_3_6_4	30.8	69.2
2_3_6_4_5	27.3	72.7
2_3_6_4_5_1	30.3	69.7

VIDEO 10_6		
	Correctly detected %	Incorrectly detected %
2	42.9	57.1
2_3	53.5	46.5
2_3_6	48.4	51.6
2_3_6_4	46.9	53.1
2_3_6_4_5	40.5	59.5
2_3_6_4_5_1	49.9	50.1

VIDEO 9_6		
	Correctly detected %	Incorrectly detected %
2	24.7	75.3
2_3	24.7	75.3
2_3_6	18.8	81.2
2_3_6_4	24.7	75.3
2_3_6_4_5	31.8	68.2
2_3_6_4_5_1	31.8	68.2

VIDEO 8_6		
	Correctly detected %	Incorrectly detected %
2	48.3	51.7
2_3	48.3	51.7
2_3_6	58.3	41.7
2_3_6_4	60.3	39.7
2_3_6_4_5	54.9	45.1
2_3_6_4_5_1	53.6	46.4

VIDEO 12_6		
	Correctly detected %	Incorrectly detected %
2	72.1	27.9
2_3	75.1	24.9
2_3_6	73.4	26.6
2_3_6_4	71.3	28.7
2_3_6_4_5	66.4	33.6
2_3_6_4_5_1	69.2	30.8

CATEGORY 7

VIDEO 7_7		
	Correctly detected %	Incorrectly detected %
2	7.8	92.2
2_3	1.9	98.1
2_3_6	35.9	64.1
2_3_6_4	22.3	77.7
2_3_6_4_5	25.3	74.7
2_3_6_4_5_1	24.3	75.7

VIDEO 11_7		
	Correctly detected %	Incorrectly detected %
2	18.3	81.7
2_3	55.1	44.9
2_3_6	44.9	55.1
2_3_6_4	39.7	60.3
2_3_6_4_5	44.8	55.2
2_3_6_4_5_1	46.3	53.7

VIDEO 10_7		
	Correctly detected %	Incorrectly detected %
2	7.1	92.9
2_3	0	100
2_3_6	52.4	47.6
2_3_6_4	57.1	42.9
2_3_6_4_5	47.6	52.4
2_3_6_4_5_1	45.2	54.8

VIDEO 9_7		
	Correctly detected %	Incorrectly detected %
2	35.4	64.6
2_3	12.2	87.8
2_3_6	25.2	74.8
2_3_6_4	24.5	75.5
2_3_6_4_5	29.3	70.7
2_3_6_4_5_1	29.9	70.1

VIDEO 8_7		
	Correctly detected %	Incorrectly detected %
2	27.5	72.5
2_3	11.4	88.6
2_3_6	52.8	47.2
2_3_6_4	51.8	48.2
2_3_6_4_5	62.2	37.8
2_3_6_4_5_1	64.2	35.8

VIDEO 12_7		
	Correctly detected %	Incorrectly detected %
2	8.9	91.1
2_3	25.9	74.1
2_3_6	24.4	75.6
2_3_6_4	19.9	80.1
2_3_6_4_5	13.9	86.1
2_3_6_4_5_1	15.9	84.1

CATEGORY 8

VIDEO 7_8		
	Correctly detected %	Incorrectly detected %
2	39.1	60.9
2_3	28.3	71.7
2_3_6	30.4	69.6
2_3_6_4	30.4	69.6
2_3_6_4_5	27.2	72.8
2_3_6_4_5_1	28.6	71.4

VIDEO 11_8		
	Correctly detected %	Incorrectly detected %
2	40.1	59.9
2_3	32.8	67.2
2_3_6	36.6	63.4
2_3_6_4	39.4	60.6
2_3_6_4_5	36.6	63.4
2_3_6_4_5_1	37.9	62.1

VIDEO 10_8		
	Correctly detected %	Incorrectly detected %
2	53.6	46.4
2_3	36.3	63.7
2_3_6	28.4	71.6
2_3_6_4	28.9	71.1
2_3_6_4_5	30.1	69.9
2_3_6_4_5_1	27.3	72.7

VIDEO 9_8		
	Correctly detected %	Incorrectly detected %
2	20.3	79.7
2_3	15.2	84.8
2_3_6	17.5	82.5
2_3_6_4	19.4	80.6
2_3_6_4_5	17.9	82.1
2_3_6_4_5_1	18.9	81.1

VIDEO 8_8		
	Correctly detected %	Incorrectly detected %
2	38.5	61.5
2_3	27.1	72.9
2_3_6	24.4	75.6
2_3_6_4	28.5	71.5
2_3_6_4_5	30.3	69.7
2_3_6_4_5_1	32.1	67.9

VIDEO 12_8		
	Correctly detected %	Incorrectly detected %
2	18.4	81.6
2_3	19.4	80.6
2_3_6	24.9	75.1
2_3_6_4	23.6	76.4
2_3_6_4_5	29.9	70.1
2_3_6_4_5_1	28.7	71.3

Table 12 - Multiclass SVM. Accumulative application of descriptors in the established order.

APPENDIX VII

VII. Improvement or declining of the performance according to the application of the new order established.

CATEGORY 1

VIDEO 7_1		
	Correctly detected %	Improve %
2	13.4	
2_3	17.5	4.1
2_3_6	29.8	0
2_3_6_4	28.9	-0.9
2_3_6_4_5	30.9	2
2_3_6_4_5_1	29.8	-1.1

VIDEO 11_1		
	Correctly detected %	Improve %
2	8.4	
2_3	40.5	32.1
2_3_6	45.6	5.1
2_3_6_4	34.3	-11.3
2_3_6_4_5	38.3	4
2_3_6_4_5_1	29.5	-8.8

VIDEO 10_1		
	Correctly detected %	Improve %
2	9.3	
2_3	6.9	-2.4
2_3_6	43.1	36.2
2_3_6_4	53.5	10.4
2_3_6_4_5	56.9	3.4
2_3_6_4_5_1	56.9	0

VIDEO 9_1		
	Correctly detected %	Improve %
2	21.2	
2_3	10.9	-10.3
2_3_6	28.4	17.5
2_3_6_4	35.5	7.1
2_3_6_4_5	37.4	1.9
2_3_6_4_5_1	37.6	0.2

VIDEO 8_1		
	Correctly detected %	Improve %
2	5.6	
2_3	6.4	0.8
2_3_6	28.8	22.4
2_3_6_4	29.6	0.8
2_3_6_4_5	33.6	4
2_3_6_4_5_1	33.6	0

VIDEO 12_1		
	Correctly detected %	Improve %
2	26.9	
2_3	55.2	28.3
2_3_6	60.9	5.7
2_3_6_4	59.8	-1.1
2_3_6_4_5	63.9	4.1
2_3_6_4_5_1	48.5	-15.4

CATEGORY 2

VIDEO 7_2		
	Correctly detected %	Improve %
2	18.7	
2_3	4.5	-14.2
2_3_6	37.1	32.6
2_3_6_4	24.4	-12.7
2_3_6_4_5	32.5	8.1
2_3_6_4_5_1	27.1	-5.4

VIDEO 9_2		
	Correctly detected %	Improve %
2	17.7	
2_3	10.4	-7.3
2_3_6	49.2	38.8
2_3_6_4	50	0.8
2_3_6_4_5	54.1	4.1
2_3_6_4_5_1	53.2	-0.9

VIDEO 11_2		
	Correctly detected %	Improve %
2	28.7	
2_3	56.3	27.6
2_3_6	56.9	0.6
2_3_6_4	64.7	7.8
2_3_6_4_5	65.3	0.6
2_3_6_4_5_1	67.1	1.8

VIDEO 8_2		
	Correctly detected %	Improve %
2	17.7	
2_3	6.7	-11
2_3_6	55.1	48.4
2_3_6_4	34.9	-20.2
2_3_6_4_5	45.7	10.8
2_3_6_4_5_1	35.9	-9.8

VIDEO 10_2		
	Correctly detected %	Improve %
2	11.7	
2_3	3.5	-8.2
2_3_6	43.5	40
2_3_6_4	45.8	2.3
2_3_6_4_5	54.1	8.3
2_3_6_4_5_1	53.2	-0.9

VIDEO 12_2		
	Correctly detected %	Improve %
2	20.1	
2_3	54.2	34.1
2_3_6	49.3	-4.9
2_3_6_4	55.5	6.2
2_3_6_4_5	61.8	6.3
2_3_6_4_5_1	57.6	-4.2

CATEGORY 3

VIDEO 7_3		
	Correctly detected %	Improve %
2	28.9	
2_3	21.9	-7
2_3_6	18.2	-3.7
2_3_6_4	31.6	13.4
2_3_6_4_5	34.4	2.8
2_3_6_4_5_1	33.7	-0.7

VIDEO 11_3		
	Correctly detected %	Improve %
2	8.1	
2_3	22.6	14.5
2_3_6	34.4	11.8
2_3_6_4	45.7	11.3
2_3_6_4_5	47.8	2.1
2_3_6_4_5_1	40.1	-7.7

VIDEO 10_3		
	Correctly detected %	Improve %
2	6.1	
2_3	12.9	6.8
2_3_6	15.4	2.5
2_3_6_4	43.5	28.1
2_3_6_4_5	47.1	3.6
2_3_6_4_5_1	39.5	-7.6

VIDEO 9_3		
	Correctly detected %	Improve %
2	9.4	
2_3	10.9	1.5
2_3_6	14.9	4
2_3_6_4	40.1	25.2
2_3_6_4_5	45.8	5.7
2_3_6_4_5_1	35.3	-10.5

VIDEO 8_3		
	Correctly detected %	Improve %
2	52.1	
2_3	22.4	-29.7
2_3_6	19.9	-2.5
2_3_6_4	66.2	46.3
2_3_6_4_5	65.6	-0.6
2_3_6_4_5_1	63.9	-1.7

VIDEO 12_3		
	Correctly detected %	Improve %
2	2.9	
2_3	13.7	10.8
2_3_6	21.8	8.1
2_3_6_4	64.5	42.7
2_3_6_4_5	65.8	1.3
2_3_6_4_5_1	62.8	-3

CATEGORY 4

VIDEO 7_4		
	Correctly detected %	Improve %
2	5.1	
2_3	2.04	-3.06
2_3_6	12.2	10.16
2_3_6_4	12.2	0
2_3_6_4_5	14.3	2.1
2_3_6_4_5_1	18.4	4.1

VIDEO 9_4		
	Correctly detected %	Improve %
2	0	
2_3	55.1	55.1
2_3_6	66.6	11.5
2_3_6_4	59.9	-6.7
2_3_6_4_5	61.8	1.9
2_3_6_4_5_1	62.2	0.4

VIDEO 11_4		
	Correctly detected %	Improve %
2	0	
2_3	7.9	7.9
2_3_6	5.4	-2.5
2_3_6_4	3.1	-2.3
2_3_6_4_5	2.7	-0.4
2_3_6_4_5_1	2.5	-0.2

VIDEO 8_4		
	Correctly detected %	Improve %
2	15.9	
2_3	2.3	-13.6
2_3_6	40.9	38.6
2_3_6_4	38.6	-2.3
2_3_6_4_5	40.9	2.3
2_3_6_4_5_1	20.5	-20.4

VIDEO 10_4		
	Correctly detected %	Improve %
2	0	
2_3	0	0
2_3_6	2.8	2.8
2_3_6_4	2.3	-0.5
2_3_6_4_5	2.1	-0.2
2_3_6_4_5_1	2.3	0.2

VIDEO 12_4		
	Correctly detected %	Improve %
2	0	
2_3	20.8	20.8
2_3_6	32.5	11.7
2_3_6_4	25.1	-7.4
2_3_6_4_5	26.5	1.4
2_3_6_4_5_1	31.8	5.3

CATEGORY 5

VIDEO 7_5		
	Correctly detected %	Improve %
2	77.7	
2_3	84.1	6.4
2_3_6	38.6	-45.5
2_3_6_4	25.8	-12.8
2_3_6_4_5	22.3	-3.5
2_3_6_4_5_1	16.5	-5.8

VIDEO 11_5		
	Correctly detected %	Improve %
2	92.3	
2_3	9.1	-83.2
2_3_6	6.5	-2.6
2_3_6_4	2.6	-3.9
2_3_6_4_5	3.5	0.9
2_3_6_4_5_1	5.1	1.6

VIDEO 10_5		
	Correctly detected %	Improve %
2	74.3	
2_3	81.7	7.4
2_3_6	40.6	-41.1
2_3_6_4	28.6	-12
2_3_6_4_5	18.7	-9.9
2_3_6_4_5_1	20.3	1.6

VIDEO 9_5		
	Correctly detected %	Improve %
2	36.8	
2_3	59.9	23.1
2_3_6	9.6	-50.3
2_3_6_4	8.5	-1.1
2_3_6_4_5	8.2	-0.3
2_3_6_4_5_1	9.6	1.4

VIDEO 8_5		
	Correctly detected %	Improve %
2	82.4	
2_3	90.4	8
2_3_6	27.8	-62.6
2_3_6_4	27.3	-0.5
2_3_6_4_5	19.8	-7.5
2_3_6_4_5_1	20.9	1.1

VIDEO 12_5		
	Correctly detected %	Improve %
2	56.7	
2_3	18	-38.7
2_3_6	29	11
2_3_6_4	36	7
2_3_6_4_5	35.7	-0.3
2_3_6_4_5_1	35.3	-0.4

CATEGORY 6

VIDEO 7_6		
	Correctly detected %	Improve %
2	34.9	
2_3	44.3	9.4
2_3_6	52.5	8.2
2_3_6_4	48.1	-4.4
2_3_6_4_5	48.1	0
2_3_6_4_5_1	49.7	1.6

VIDEO 9_6		
	Correctly detected %	Improve %
2	24.7	
2_3	24.7	0
2_3_6	18.8	-5.9
2_3_6_4	24.7	5.9
2_3_6_4_5	31.8	7.1
2_3_6_4_5_1	31.8	0

VIDEO 11_6		
	Correctly detected %	Improve %
2	36.3	
2_3	37.8	1.5
2_3_6	32.1	-5.7
2_3_6_4	30.8	-1.3
2_3_6_4_5	27.3	-3.5
2_3_6_4_5_1	30.3	3

VIDEO 8_6		
	Correctly detected %	Improve %
2	48.3	
2_3	48.3	0
2_3_6	58.3	10
2_3_6_4	60.3	2
2_3_6_4_5	54.9	-5.4
2_3_6_4_5_1	53.6	-1.3

VIDEO 10_6		
	Correctly detected %	Improve %
2	42.9	
2_3	53.5	10.6
2_3_6	48.4	-5.1
2_3_6_4	46.9	-1.5
2_3_6_4_5	40.5	-6.4
2_3_6_4_5_1	49.9	9.4

VIDEO 12_6		
	Correctly detected %	Improve %
2	72.1	
2_3	75.1	3
2_3_6	73.4	-1.7
2_3_6_4	71.3	-2.1
2_3_6_4_5	66.4	-4.9
2_3_6_4_5_1	69.2	2.8

CATEGORY 7

VIDEO 7_7		
	Correctly detected %	Improve %
2	7.8	
2_3	1.9	-5.9
2_3_6	35.9	34
2_3_6_4	22.3	-13.6
2_3_6_4_5	25.3	3
2_3_6_4_5_1	24.3	-1

VIDEO 11_7		
	Correctly detected %	Improve %
2	18.3	
2_3	55.1	36.8
2_3_6	44.9	-10.2
2_3_6_4	39.7	-5.2
2_3_6_4_5	44.8	5.1
2_3_6_4_5_1	46.3	1.5

VIDEO 10_7		
	Correctly detected %	Improve %
2	7.1	
2_3	0	-7.1
2_3_6	52.4	52.4
2_3_6_4	57.1	4.7
2_3_6_4_5	47.6	-9.5
2_3_6_4_5_1	45.2	-2.4

VIDEO 9_7		
	Correctly detected %	Improve %
2	35.4	
2_3	12.2	-23.2
2_3_6	25.2	13
2_3_6_4	24.5	-0.7
2_3_6_4_5	29.3	4.8
2_3_6_4_5_1	29.9	0.6

VIDEO 8_7		
	Correctly detected %	Improve %
2	27.5	
2_3	11.4	-16.1
2_3_6	52.8	41.4
2_3_6_4	51.8	-1
2_3_6_4_5	62.2	10.4
2_3_6_4_5_1	64.2	2

VIDEO 12_7		
	Correctly detected %	Improve %
2	8.9	
2_3	25.9	17
2_3_6	24.4	-1.5
2_3_6_4	19.9	-4.5
2_3_6_4_5	13.9	-6
2_3_6_4_5_1	15.9	2

CATEGORY 8

VIDEO 7_8		
	Correctly detected %	Improve %
2	39.1	
2_3	28.3	-10.8
2_3_6	30.4	2.1
2_3_6_4	30.4	0
2_3_6_4_5	27.2	-3.2
2_3_6_4_5_1	28.6	1.4

VIDEO 9_8		
	Correctly detected %	Improve %
2	20.3	
2_3	15.2	-5.1
2_3_6	17.5	2.3
2_3_6_4	19.4	1.9
2_3_6_4_5	17.9	-1.5
2_3_6_4_5_1	18.9	1

VIDEO 11_8		
	Correctly detected %	Improve %
2	40.1	
2_3	32.8	-7.3
2_3_6	36.6	3.8
2_3_6_4	39.4	2.8
2_3_6_4_5	36.6	-2.8
2_3_6_4_5_1	37.9	1.3

VIDEO 8_8		
	Correctly detected %	Improve %
2	38.5	
2_3	27.1	-11.4
2_3_6	24.4	-2.7
2_3_6_4	28.5	4.1
2_3_6_4_5	30.3	1.8
2_3_6_4_5_1	32.1	1.8

VIDEO 10_8		
	Correctly detected %	Improve %
2	53.6	
2_3	36.3	-17.3
2_3_6	28.4	-7.9
2_3_6_4	28.9	0.5
2_3_6_4_5	30.1	1.2
2_3_6_4_5_1	27.3	-2.8

VIDEO 12_8		
	Correctly detected %	Improve %
2	18.4	
2_3	19.4	1
2_3_6	24.9	5.5
2_3_6_4	23.6	-1.3
2_3_6_4_5	29.9	6.3
2_3_6_4_5_1	28.7	-1.2

Table 13 - Improvement or declining of the performance according to the application of the new order established

APPENDIX VIII

VIII. a) Multiclass SVM. Addition of the new descriptor: Dominant Color

Table 14 - Multiclass SVM. Addition of the new descriptor: Dominant Color

CATEGORY 1				
Video	All features		Dominant Color Extractor	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_1	56.8	43.2	58.4	41.6
Video_8_1	28.4	71.6	0	100
Video_9_1	33.7	66.3	0	100
Video_10_1	49.6	50.4	88.5	11.5
Video_11_1	46.9	53.1	78.3	21.7
Video_12_1	60.4	39.6	52.6	47.4

CATEGORY 2				
Video	All features		Dominant Color Extractor	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_2	65.3	34.7	0	100
Video_8_2	55.6	44.4	0	100
Video_9_2	28.2	71.8	0	100
Video_10_2	42.5	57.5	0	100
Video_11_2	71.3	28.7	0	100
Video_12_2	66.7	33.3	0	100

CATEGORY 3				
Video	All features		Dominant Color Extractor	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_3	19.8	80.2	21.4	78.6
Video_8_3	60.4	39.6	0	100
Video_9_3	15.2	84.8	59.3	40.7
Video_10_3	12.9	87.1	97.5	2.5
Video_11_3	30.1	69.9	84.1	15.9
Video_12_3	50	50	0	100

CATEGORY 4				
Video	All features		Dominant Color Extractor	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_4	28.6	71.4	94.9	5.1
Video_8_4	25	75	81.8	18.2
Video_9_4	63.3	36.7	0	100
Video_10_4	4.2	95.8	0	100
Video_11_4	3.4	96.6	0	100
Video_12_4	32.5	67.5	0	100

CATEGORY 5				
Video	All features		Dominant Color Extractor	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_5	40.6	59.4	36.9	63.1
Video_8_5	52.1	47.9	30.9	69.1
Video_9_5	0.4	99.6	0	100
Video_10_5	44.3	55.7	58.7	41.3
Video_11_5	80.5	19.5	89.3	10.7
Video_12_5	52.7	47.3	20	80

CATEGORY 6				
Video	All features		Dominant Color Extractor	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_6	51.9	48.1	22.4	77.6
Video_8_6	58.3	41.7	15.2	84.8
Video_9_6	17.6	82.4	7.1	92.9
Video_10_6	55.9	44.1	20.8	79.2
Video_11_6	30.9	69.1	17.8	82.2
Video_12_6	62.8	37.2	0	100

CATEGORY 7				
Video	All features		Dominant Color Extractor	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_7	61.7	38.3	0	100
Video_8_7	24.5	75.5	0	100
Video_9_7	57.1	42.9	64.3	35.7
Video_10_7	59.9	40.1	72.6	27.4
Video_11_7	52.8	47.2	0	100
Video_12_7	40.8	59.2	88.1	11.9

CATEGORY 8				
Video	All features		Dominant Color Extractor	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_8	18.5	81.5	0	100
Video_8_8	27.6	72.4	11.3	88.7
Video_9_8	6.9	93.1	0	100
Video_10_8	6.1	93.9	0	100
Video_11_8	8.4	91.6	0	100
Video_12_8	15.9	84.1	0	100

VIII. b) Multiclass SVM. Addition of the new descriptor: Color Layout

Table 15 - Multiclass SVM. Addition of the new descriptor: Color Layout

CATEGORY 1				
Video	All features		Color Layout	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_1	56.8	43.2	24.8	75.2
Video_8_1	28.4	71.6	18.9	81.1
Video_9_1	33.7	66.3	16.3	83.7
Video_10_1	49.6	50.4	19.4	80.6
Video_11_1	46.9	53.1	18.8	81.2
Video_12_1	60.4	39.6	21.5	78.5

CATEGORY 2				
Video	All features		Color Layout	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_2	65.3	34.7	18.2	81.8
Video_8_2	55.6	44.4	12.9	87.1
Video_9_2	28.2	71.8	13.3	86.7
Video_10_2	42.5	57.5	19.5	80.5
Video_11_2	71.3	28.7	23.9	76.1
Video_12_2	66.7	33.3	12.5	87.5

CATEGORY 3				
Video	All features		Color Layout	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_3	19.8	80.2	16.4	83.6
Video_8_3	60.4	39.6	19.6	80.4
Video_9_3	15.2	84.8	19.2	80.7
Video_10_3	12.9	87.1	15.9	84.1
Video_11_3	30.1	69.9	19.3	80.7
Video_12_3	50	50	16.2	83.8

CATEGORY 4				
Video	All features		Color Layout	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_4	28.6	71.4	22.2	77.8
Video_8_4	25	75	15.9	84.1
Video_9_4	63.3	36.7	15.7	84.3
Video_10_4	4.2	95.8	14.6	85.4
Video_11_4	3.4	96.6	18.3	81.7
Video_12_4	32.5	67.5	21.2	78.8

CATEGORY 5				
Video	All features		Color Layout	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_5	40.6	59.4	13.7	86.3
Video_8_5	52.1	47.9	18.5	81.5
Video_9_5	0.4	99.6	17.5	82.5
Video_10_5	44.3	55.7	16.6	83.4
Video_11_5	80.5	19.5	17.2	82.8
Video_12_5	52.7	47.3	16.8	83.2

CATEGORY 6				
Video	All features		Color Layout	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_6	51.9	48.1	12.9	87.1
Video_8_6	58.3	41.7	17.8	82.2
Video_9_6	17.6	82.4	12.4	87.6
Video_10_6	55.9	44.1	13.2	86.8
Video_11_6	30.9	69.1	12.4	87.6
Video_12_6	62.8	37.2	14.4	85.6

CATEGORY 7				
Video	All features		Color Layout	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_7	61.7	38.3	19.1	80.9
Video_8_7	24.5	75.5	20.4	79.6
Video_9_7	57.1	42.9	23.3	76.7
Video_10_7	59.9	40.1	19.9	80.1
Video_11_7	52.8	47.2	23.2	76.8
Video_12_7	40.8	59.2	23.3	76.7

CATEGORY 8				
Video	All features		Color Layout	
	Correctly detected %	Incorrectly detected %	Correctly detected %	Incorrectly detected %
Video_7_8	18.5	81.5	16.3	83.7
Video_8_8	27.6	72.4	18.3	81.7
Video_9_8	6.9	93.1	17.7	82.3
Video_10_8	6.1	93.9	20.5	79.5
Video_11_8	8.4	91.6	18.1	81.9
Video_12_8	15.9	84.1	18.7	81.3

APPENDIX C. Presupuesto

1) Ejecución Material

- Compra de ordenador personal (Software incluido)..... ..... 2.000 €
- Alquiler de impresora láser durante 6 meses ..... 260 €
- Material de oficina ..... 150 €
- Total de ejecución material..... 2.400 €

2) Gastos generales

- 16 % sobre Ejecución Material..... 352 €

3) Beneficio Industrial

- 6 % sobre Ejecución Material..... 132 €

4) Honorarios Proyecto

- 1800 horas a 15 € / hora..... 27.000 €

5) Material fungible

- Gastos de impresión ..... 280 €
- Encuadernación ..... 200 €

6) Subtotal del presupuesto

- Subtotal Presupuesto..... 32.774 €

7) I.V.A. aplicable

- 16% Subtotal Presupuesto..... 5.899,3 €

8) Total presupuesto

- Total Presupuesto ..... 38.673,8 €

Madrid, Septiembre 2011

El Ingeniero Jefe de Proyecto

Fdo.: Loreto Felipe Sánchez-Infante
Ingeniero Superior de Telecomunicación

APPENDIX D. Pliego de Condiciones

PLIEGO DE CONDICIONES

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un sistema basado en la *clasificación automática de vídeo por género*. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad “Presupuesto de Ejecución de Contrata” y anteriormente llamado ”Presupuesto de Ejecución Material” que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.