

# Descoberta de gêneros de filmes utilizando mineração de vídeo

André Alex Sestari<sup>1</sup>, Antônio Marco da Silva<sup>1</sup>, Tiago P. da Silva<sup>1</sup>

<sup>1</sup>Departamento de Sistemas e Computação  
Universidade Regional de Blumenau (FURB) – Blumenau, SC – Brazil

andre.sestari@hotmail.com, antomarsi@hotmail.com,  
tiagops.prog@gmail.com

**Abstract.** *This paper describes the tools, project structure, methodology and the results achieved applying video mining to discover movies genres considering the frame colors and its variations.*

**Resumo.** *Este artigo descreve as ferramentas, estrutura de projeto, metodologia e o resultado alcançado ao aplicar a mineração de vídeo para descoberta de gêneros de filmes considerando as cores dos quadros do filme e suas variações.*

## 1. Informações Gerais

Abraçando a proposta do professor Claudio Ratke de estender um trabalho anterior que utilizava as cores de quadros para obtenção do gênero, foi implementada uma versão similar que além das cores de pontos específicos de cada cena considera a variação das cores obtidas em cada quadro.

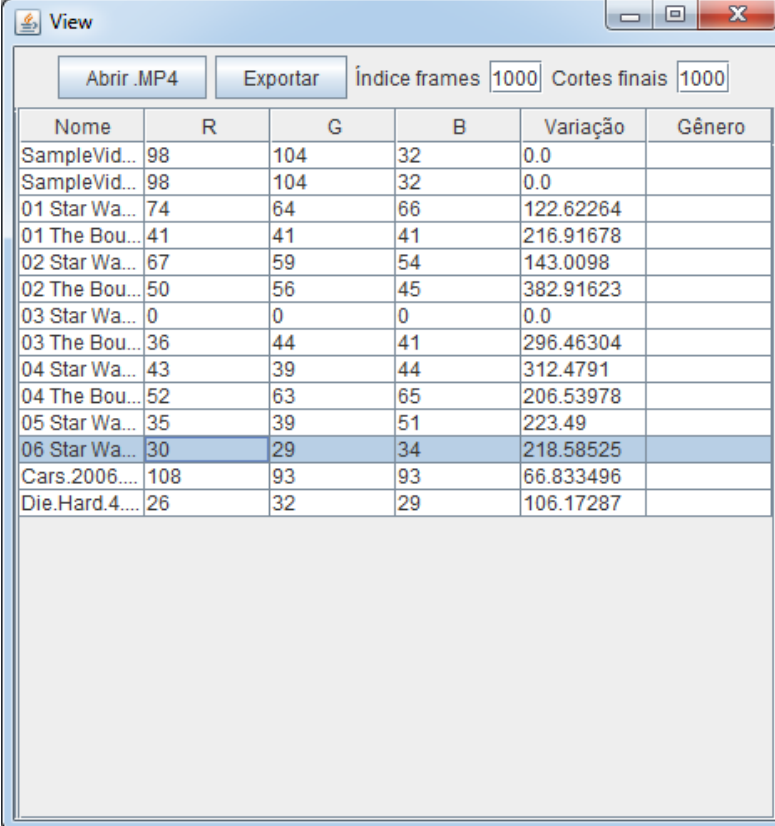
## 2. Estrutura do projeto

A aplicação foi desenvolvida utilizando a linguagem Java, tendo como foco a simplicidade no desenvolvimento e a facilidade de manipulação.

Utilizado Maven para controle de dependências, IDE Eclipse para o desenvolvimento e GitHub para controle de versão.

Composta por duas classes, View e VideoRGB, tratamos o recebimento do arquivos em uma e guardamos a referência para as cores e o cálculo de variação em outra, resultando em duas classes de médio porte.

Para a tela, temos um grid principal onde exibimos os dados coletados no processamento de cada filme, com os botões Abrir, onde selecionamos o filme, Exportar, onde exportamos o grid para excel, Índice frames, onde definimos a proporção de frames lidos, e Corte, onde definimos quantos cortes serão feitos na lista final de cores de cada filme.



Nome	R	G	B	Variação	Gênero
SampleVid...	98	104	32	0.0	
SampleVid...	98	104	32	0.0	
01 Star Wa...	74	64	66	122.62264	
01 The Bou...	41	41	41	216.91678	
02 Star Wa...	67	59	54	143.0098	
02 The Bou...	50	56	45	382.91623	
03 Star Wa...	0	0	0	0.0	
03 The Bou...	36	44	41	296.46304	
04 Star Wa...	43	39	44	312.4791	
04 The Bou...	52	63	65	206.53978	
05 Star Wa...	35	39	51	223.49	
06 Star Wa...	30	29	34	218.58525	
Cars.2006....	108	93	93	66.833496	
Die.Hard.4....	26	32	29	106.17287	

**Figure 1. Tela da aplicação**

## 2. Obtenção das cores

Utilizando uma ferramenta chamada bytedeco, carregamos o filme no formato mp4 e percorremos todos os seus frames, guardando quatro RGBs dos cantos de cada imagem.

Por medidas de performance, limitamos a quantidade de frames capturados para um em cada mil. Os filmes utilizados possuem uma taxa de aproximadamente 25 frames por segundo, resultando em 135 mil frames para um cada 1 hora e meia de vídeo.

Para realizar a mineração com filmes/vídeos mais curtos, devemos alterar essa taxa de frames para termos uma quantidade de valores útil.

### 2.1. Média

Após a obtenção das cores, passamos a lista para uma classe auxiliar que irá realizar um novo corte nos frames para o cálculo da média e variação. Esse corte é feito em 10 seções, sendo que as duas últimas são ignoradas por conta dos créditos.

Como esse corte é feito de acordo com a quantidade de frames lidos (um filme com 70 minutos terá 100 frames lidos e cada seção terá um intervalo de 7 minutos), em filmes mais longos perderemos o final no cálculo.

Podemos perceber uma grande variação de assertividade de acordo com o número de cortes e o número de frames.

### **3. Taxa de variação**

Junto com o cálculo da média de cores das seções, realizamos um cálculo de variação. Esse cálculo é feito com a média das médias da diferença das cores equivalentes de seções adjacentes dividida pela cor da seção de menor ordem entre elas multiplicada por 100. Caso o valor seja negativo, multiplicamos por -1 para ignorar a ordem da variação.

Com esse cálculo garantimos que a variação seja calculada de uma seção para sua sucessora também considerando a variação de todo o filme.

### **4. Mineração**

Após a obtenção dos valores em excel, informamos os gêneros de cada filme manualmente em uma coluna “Gênero”.

Passando a tabela para o RapidMiner, geramos uma árvore de decisão tendo como label o campo “Gênero”, o que resultou em uma árvore bem expressiva.

#### **4.1. Árvore de decisão**

Realizamos a mineração com 54 filmes de base, sabendo que é um espectro pequeno para um cálculo assertivo. Utilizando o RapidMiner, criamos uma árvore de decisão sem profundidade máxima.

A árvore resultado possui 21 folhas e 20 nós. 4 folhas de animação, 3 de romance, 3 de terror, 8 de aventura e 3 de ação. Concluimos que a diferença no número de folhas com resultado ação se dá pelo input de filmes (maioria aventura), e pela categoria em si. Aventura é um dos gêneros mais abrangentes para filmes, podendo variar muito de um filme para outro.

### **4. Resultado**

Descobrimos que a taxa de variação não representou grandes alterações na árvore de decisão. A maior diferença na assertividade se percebeu ao alterar a quantidade de frames lidos, mas o custo dessa busca de frames grande para a CPU é muito alto.

### **Referências**

Claudio Ratke. (2016) “Árvore de decisão”.

Claudio Ratke. (2016) “Mineração de processos”.

Bytedeco, múltiplos autores (2016) “Java interface to OpenCV and more”, <https://github.com/bytedeco/javacv>, December.