

## 5 Resultados e testes

Este capítulo apresenta os resultados obtidos pelos algoritmos desenvolvidos para classificação e segmentação. Estes resultados foram obtidos através de testes realizados com um acervo de vídeos de diferentes tipos, para que os algoritmos fossem testados de forma o mais completa possível. Esses vídeos foram obtidos em acervos digitais gratuitos, disponíveis na Internet. Os detalhes do acervo analisado e os resultados obtidos em cada algoritmo são detalhados a seguir.

### 5.1. Acervo de vídeos

Na seleção de vídeos para análise, procurou-se obter vídeos de tipos variados, de forma a testar a eficiência dos algoritmos com diferentes tipos de vídeos. Foram selecionados vários vídeos com duração total de mais de uma hora, os quais foram analisados pelos dois algoritmos. Há vídeos que possuem um tipo único de áudio, vídeos que possuem somente fala, e vídeos onde todos as diferentes classificações se misturam. Os vídeos analisados dividem-se entre os tipos:

- Clipes de música – nesses vídeos há somente música.
- Reportagens de telejornal – nesses vídeos há somente a fala, com a presença de ruídos de fundo no caso de reportagens fora do estúdio.
- Seriados de televisão – temos aqui a presença de todos os tipos de classificações, incluindo aplausos vindos da claque presente na maioria dos seriados de comédia.
- Apresentações de música ao vivo em programas de entrevista – são caracterizados por haver uma introdução da banda pelo apresentador do programa, com aplausos do público e um fechamento novamente pelo apresentador, além da performance da banda. Com isso, nesses vídeos há a presença de todos os tipos de classificações.

- Entrevistas em programas ao vivo – nesse caso, há sempre a presença de fala e aplausos, além de música nos casos onde a banda do programa faz uma intervenção na apresentação ou despedida do convidado.

- Comerciais de televisão – os tipos de áudio encontrados nesses vídeos variam.

Com esses tipos de vídeos, podemos considerar que os testes demonstram a versatilidade dos algoritmos e sua capacidade de trabalhar com qualquer tipo de vídeo apresentado.

## 5.2. Classificação

O algoritmo de classificação demonstrou nos testes um ótimo resultado, comparado à análise humana, feita a partir da audição dos vídeos e anotação dos tipos de áudio encontrados em cada intervalo. Para a maior parte dos tipos de vídeo e áudio, a média de acertos ficou acima de 92%. A tabela abaixo ilustra a média de acertos obtida por tipo de vídeo.

<b>Descrição</b>	<b>Porcentagem de acertos</b>
Clipes de música	92,5 %
Reportagens de telejornal	95 %
Seriados de televisão	92 %
Música ao vivo em programas de entrevistas	97,5 %
Entrevistas em programas ao vivo	94 %
Comerciais de televisão	92 %

Tabela 1 – Média de acertos da classificação por tipo de vídeo

A próxima tabela ilustra os resultados da classificação a partir do tipo de áudio contido nos vídeos analisados.

<b>Descrição</b>	<b>Porcentagem de acertos</b>
Música pop/rock	97 %
Música eletrônica	92 %
Música clássica	88 %
Somente fala, com e sem ruídos	96 %
Fala e aplausos	93 %
Fala e música	94 %
Todos os tipos	96 %

Tabela 2 – Média de acertos da classificação por tipo de áudio

Os resultados, quando comparados aos do trabalho no qual a idéia inicial da classificação se baseia, comprovam que houve uma evolução, mesmo com a adição da detecção de aplausos, que aumenta a abrangência da classificação. Nesse trabalho, \*\*\* atingem uma taxa de acertos média de 92%, enquanto o algoritmo proposto nesta dissertação tem uma média de acertos total de aproximadamente 94%. As taxas de acertos encontradas se equivalem às de estudos que trabalham com uma quantidade maior de informações, como os valores de amostras ou até a onda sonora reconstruída, o que demonstra sua maior eficácia.

Pode-se observar que o pior resultado encontrado é na análise de música clássica. Esse foi o tipo de áudio com os piores resultados iniciais, e por isso foi desenvolvida uma técnica específica, referenciada no capítulo 3, visando melhorar o desempenho do algoritmo nesse caso, que aumentou sua precisão para o nível apresentado, que já é bem satisfatório.

Outro caso onde o resultado inicial foi fraco foi em música eletrônica. Devido às suas características, o resultado da análise muitas vezes indicava fala, e não música. Como demonstrado, o desenvolvimento da análise rítmica resolveu

esse problema, e o método pôde atingir uma porcentagem alta de acertos também nessa análise.

Esses casos são mencionados para demonstrar a evolução do algoritmo a partir das novas idéias apresentadas. Na parte de trabalhos futuros, é sugerida a incorporação de técnicas adicionais, que podem resolver problemas específicos.

### **5.3. Segmentação**

Para a análise dos resultados da segmentação, primeiro é necessário fazer uma definição de termos e métodos de análise. Em seguida, os resultados atingidos são expostos e analisados.

#### **5.3.1. Métodos e termos**

Para medir o desempenho do algoritmo apresentado nesta pesquisa, são usadas três medidas: a taxa de acertos, a taxa de cortes não detectados e a taxa de cortes falsos detectados.

A taxa de acertos estabelece a razão entre o número de mudanças de cena encontradas e o número real de mudanças de cena (Equação 1). A taxa de cortes não detectados, considerada taxa de perda, estabelece a razão entre o número de mudanças de cena não detectadas e o número real de mudanças de cena (Equação 2). Já a terceira medida utilizada, a taxa de cortes falsos, estabelece a razão entre o número total de falsas mudanças de cena detectadas e o número real de mudanças de cena do vídeo (Equação 3).

$$\text{Taxa de acertos} = \frac{\text{total de cortes corretos detectados}}{\text{total de cortes reais}} \quad (1)$$

$$\text{Taxa de perda} = \frac{\text{total de cortes não detectados}}{\text{total de cortes reais}} \quad (2)$$

$$\text{Taxa de cortes falsos} = \frac{\text{total de cortes falsos detectados}}{\text{total de cortes reais}} \quad (3)$$

### 5.3.2. Resultados

O método desenvolvido para a segmentação do áudio mostrou como característica principal a alta taxa de acertos e a baixa taxa de perda, porém com uma alta taxa de cortes falsos para alguns dos tipos de vídeos que foram analisados.

Assim, primeiramente os resultados encontrados vão ser expostos, e em seguida será feita uma análise desses resultados, explicando o potencial do algoritmo proposto e identificando situações onde ainda há possibilidade de melhora em seu desempenho.

A Tabela 4 mostra o desempenho do algoritmo a partir do tipo de vídeo analisado. Vale observar que a categoria “comerciais de televisão” não aparece nessa tabela, ao contrário do que ocorre na classificação, já que os comerciais analisados eram únicos, e não tinham mudanças de cena. Sendo assim, eles foram avaliados apenas em relação à classificação.

<b>Descrição</b>	<b>Taxa de acertos</b>	<b>Taxa de perda</b>	<b>Taxa de cortes falsos</b>
Clipes de música	0,93	0,07	0,47
Reportagens de telejornal	0,94	0,06	3,00
Seriados de televisão	0,92	0,08	1,60
Música ao vivo em programas de entrevistas	0,89	0,11	0,67
Entrevistas em programas ao vivo	1,00	0,00	1,50

Tabela 3 – Desempenho do algoritmo de segmentação a partir do tipo de vídeo

Como se pode ver, a taxa de acertos é muito alta para todos os tipos de vídeos analisados. Isso demonstra que a parte de detecção de cortes do algoritmo é bastante eficiente com os parâmetros utilizados. Um relaxamento do limite no grau de novidade, por exemplo, poderia aumentar ainda mais essa taxa de acertos, porém isso em consequência acarretaria um aumento da taxa de cortes falsos, o que deve ser evitado.

Outro resultado evidente é que a taxa de cortes falsos para vídeos onde a fala é predominante ainda é muito alta. Isso é causado por diversos fatores. Um deles é a variação do modo de falar de cada interlocutor. Por exemplo, línguas diferentes (presentes nas amostras de vídeos analisadas) formam padrões diferentes na fala, o que dificulta o desenvolvimento de um filtro a partir apenas dos padrões de pausa. A fala tem como característica surtos de voz entremeados por intervalos de silêncio pequenos ou longos, dependendo se há o fim de uma palavra, sentença, ou uma linha de pensamento do interlocutor e do próprio ritmo da fala.

No capítulo 7 desta dissertação, será mencionado o desenvolvimento de novos filtros para a análise dos candidatos a mudanças de cena que ocorrem durante a fala. Com esse desenvolvimento, espera-se que seja possível diminuir essa taxa de cortes falsos, e tornar o algoritmo proposto ainda mais eficiente como uma alternativa para a segmentação digital de áudio. Outro fator mencionado é a utilização das informações da imagem conjuntamente à do áudio. Em vídeos onde a fala é predominante, é muito comum que as mudanças ocorram simultaneamente

entre a imagem e o áudio. Assim, a inclusão dessas informações na análise certamente possibilitaria uma maior precisão para eliminação de cortes falsos.

Uma opção de filtro observada nos testes foi a possibilidade de um ranqueamento dos candidatos a cortes a partir do grau de novidade. Em alguns casos onde a taxa de cortes falsos foi muito alta, esse ranqueamento indicou os cortes reais como os principais candidatos a mudanças de cena. Por falta de tempo, esse filtro não foi desenvolvido em detalhes e incluído no método, porém será um fator observado na parte relativa aos trabalhos futuros, e é mais um indicativo do potencial do algoritmo sugerido nesta dissertação.

É importante mencionar que não apenas o algoritmo tem variações de comportamento a partir do tipo de vídeo, mas também os tipos de cortes considerados variam. A seleção dos cortes reais de cada vídeo foi feita a partir do trabalho humano, assistindo os vídeos e marcando momentos de corte. Assim, quais eventos indicam cortes devem ser mencionados.

O primeiro tipo de corte considerado se baseia no próprio conceito de cena, típico de filmes e seriados, por exemplo. Nesses casos, são consideradas mudanças de cena não somente a troca de cena clássica, com mudança de contexto e personagens, mas também segmentos dentro de uma mesma cena. Um exemplo disso seria uma cena ocorrendo dentro de um apartamento, onde há dois diferentes grupos de personagens dialogando. A troca de foco de um grupo para outro é considerada uma mudança de cena nesta dissertação.

Outro tipo de corte considerado é o corte a partir da mudança de classificação do áudio. Nele, qualquer mudança do tipo de áudio que está ocorrendo é considerada uma mudança de cena. Sendo assim, intervenções de aplausos, por exemplo, são consideradas um segmento. Outro segmento seria uma intervenção musical no meio de segmentos de fala. Em qualquer momento que a classificação muda, considera-se que há um novo segmento.

Um caso específico deve ser mencionado para vídeos de música. Alguns dos vídeos analisados possuíam apenas um clipe de música, ou uma apresentação musical, sem que houvesse um indicativo forte de mudança de cena, pelo áudio conter somente a música apresentada. Nesses casos, as mudanças fortes de dinâmica da música são os cortes indicados. Exemplos disso são a entrada de vocais e guitarra após uma introdução, o começo de um refrão onde há um

aumento significativo na dinâmica da música, ou a interrupção da música para um solo de guitarra.

A abrangência dos tipos de cortes considerados mostra mais uma vez o potencial do algoritmo desenvolvido. São percebidas mudanças até certo ponto sutis, como a entrada de um refrão em uma música, e que não podem ser consideradas na análise da imagem. Com isso, pode-se perceber a importância da análise do áudio para a segmentação. Ao utilizar a análise do áudio como complemento à da imagem, pode-se obter resultados precisos.