

## **AVALIAÇÃO EXPLORATÓRIA PARA DETECÇÃO DE VÍDEOS DO TIPO *STUFFING***

**Albano Negreiros<sup>1</sup>, Angelo N. G. Santos<sup>1</sup>, Hugo S. Silva<sup>1</sup>, Antonio da Luz Jr.<sup>2</sup>**

<sup>1</sup>Aluno do CST em GTI - Campus Paraíso do Tocantins - IFTO. e-mail: <albanonegreiros; angelo.neto333; huggomano@gmail.com>

<sup>2</sup>Núcleo de Informação e Comunicação - Campus Paraíso do Tocantins - IFTO. e-mail: <daluz@ifto.edu.br>

**Resumo:** Filtragem de conteúdo em redes sociais é uma tarefa em constante crescimento, devido entre outras coisas, ao fato de que com o aumento da popularidade dessas redes o número de abusos tende a crescer, irritando usuários e prejudicando os serviços oferecidos. Neste trabalho estamos focados na detecção de vídeos não-colaborativos em redes sociais para compartilhamento de vídeo. Especificamente, investigamos o quanto a análise baseada em conteúdo visual pode ajudar a detectar vídeos do tipo *stuffing* em *threads* de vídeo-resposta. Esta é uma tarefa bastante desafiadora, devido o alto-nível dos conceitos semânticos envolvidos, a natureza variada das redes sociais impede o uso de conhecimento prévio acerca dos comportamentos e, um dos pontos mais importantes, a natureza de dependência ao contexto na caracterização de vídeos não-colaborativos. Investigamos qual o desempenho de diferentes estratégias de descrição de conteúdo visual e qual o impacto do uso de diferentes configurações nos parâmetros do classificador SVM na resolução desse tipo de problema.

**Palavras-chave:** cbir, cbvir, stip, svm, videos não-colaborativos

### **1 INTRODUÇÃO**

As Redes Sociais Virtuais (Newman e Park, 2003), dentre os novos modos de comunicação gerados com a massificação do acesso em alta velocidade à Internet, são dos que têm conseguido maior volume de usuários, em um conjunto impressionante de aplicações. Associado a isso, o avanço da tecnologia de geração de vídeo digital, permitiu a criação de uma nova forma de interação entre as pessoas, as Redes Sociais para Compartilhamento de Vídeo Digital (Luz Jr., 2012), como por exemplo, o Youtube.

Os usuários dessas redes estão suscetíveis a ações positivas e negativas de outros membros. Os possíveis abusos e infrações cometidas pelos usuários são entendidas como não-colaborativas, pois poluem o canal de comunicação e não contribuem para o alcance dos objetivos originais das comunidades, aprendizado e entretenimento. Entre essas ações estão: instigar a irritação de outros usuários, difundir material inapropriado para os fins da comunidade (por ex., propagandas ou pornografia) e manipular ilegitimamente a lista de classificação de popularidade (Luz Jr., 2012).

Conforme observado por Backstrom et al. (2006), a permanência dos usuários em redes sociais tem estreita relação com o modo como são tratados os seus tópicos de interesse. E segundo Luz Jr. (2012), além de implicar a redução da satisfação com o serviço, a ação não-colaborativa consome recursos computacionais, que certamente seriam mais bem aproveitados se empregados para fins construtivos.

Um exemplo de ação não-colaborativa é a identificação da presença de vídeos-resposta que possuem conteúdo com semântica divergente ao tema do vídeo original a que foi associado. Segundo Luz Jr. (2012), esse tipo de ação pode ser agrupada em duas classes distintas, de acordo com as suas características: *stuffing* e *spam*. Neste trabalho foi abordado apenas o tipo *stuffing*. A Figura 1 apresenta o vídeo original, respostas legítimas e exemplos de vídeos *stuffing*.

Neste trabalho tivemos como objetivo realizar uma investigação quanto ao desempenho de diferentes estratégias de descrição de conteúdo visual, tais como: Histograma de Cores RGB (Gonzalez e Woods, 2006), Momentos de Zernike (Khotanzad e Hong, 1990), SIFT (Lowe, 1999) e STIP (Laptev, 2005). Ainda, foi verificado como cada uma dentre diferentes configurações de parâmetros de *kernel* do classificador SVM (Vapnik, 1979) impacta na busca por uma solução ótima para o problema de identificação de vídeos do tipo *stuffing*.

## 2 METODOLOGIA

O presente trabalho tem por característica a pesquisa quanti-qualitativa, bibliográfica e exploratória. Foi fundamentado na metodologia elaborada por Luz Jr. (2012) e adaptado para possibilitar a formação de pessoal qualificado na área de recuperação de imagens e vídeos com base no conteúdo, permitindo a participação de alunos de Iniciação Científica e Iniciação Científica Júnior.

Após etapa inicial de formação teórica acerca das áreas de Aprendizado de Máquina, Recuperação de Informação com Base no Conteúdo e Redes Sociais, foram realizadas as etapas de construção da base de dados, anotação manual de vídeos, implementação e execução de algoritmos para extração e representação de características visuais e avaliação de resultados.

## 3 RESULTADOS E DISCUSSÕES

### 3.1. Análise comparativa do desempenho de descritores visuais

Nesta seção são apresentados os resultados obtidos nos experimentos de identificação da presença de vídeos do tipo *stuffing* em *threads* de vídeo-resposta, utilizando diferentes categorias de descritores visuais. Os algoritmos avaliados foram definidos após o estudo de algumas das classes de técnicas existentes, sendo selecionados os mais significativos dentre abordagens variadas: a) Histograma nos canais RGB, um algoritmo de âmbito global, utilizado para descrever o comportamento da distribuição de cores em uma imagem; b) Momentos de Zernike, também de âmbito global, descreve um conjunto de Momentos Invariantes associados a informação de textura

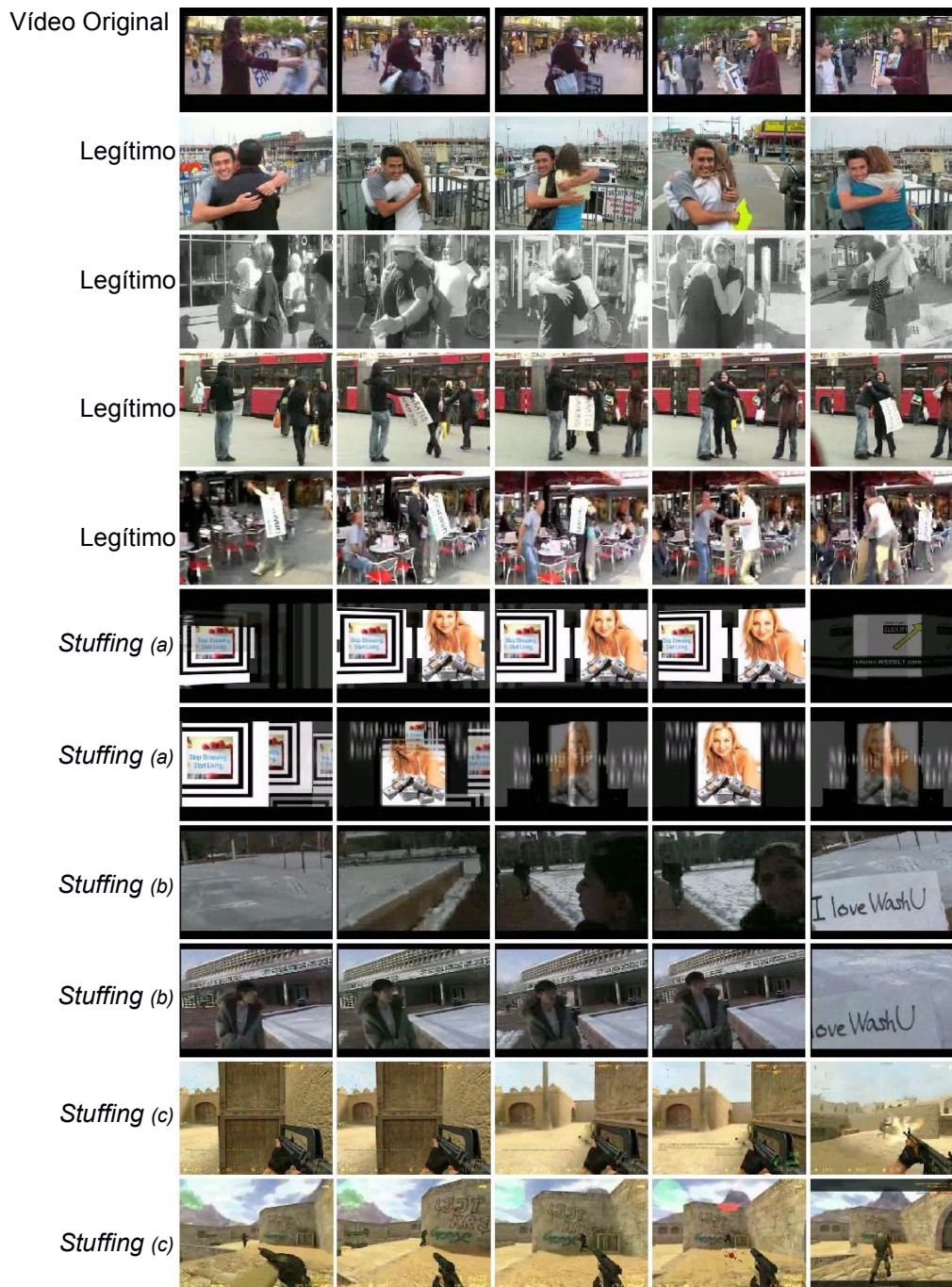


Figura 1. Quadros extraídos de vídeos exemplificando a ocorrência de elementos da categoria *Stuffing*. Os vídeos que compartilham conteúdo estão identificados com uma letra entre parênteses. Fonte: (Luz Jr., 2012).

dos objetos presentes na imagem; c) SIFT, algoritmo de âmbito local, identifica e descreve um conjunto de pontos de interesse na imagem, os quais não sofrem influência de variações de cor, escala e orientação, e; d) STIP, algoritmo de âmbito local, que, assim como o SIFT, faz a identificação e descrição de um conjunto de pontos de interesse, contudo, diferentemente dos

demais algoritmos, este foi construído para trabalhar diretamente com vídeos e não sobre imagens extraídas destes.

Os experimentos foram realizados utilizando o ambiente Weka (Hall et al., 2009) e classificador SVM, com *kernel* Linear, e demais configurações padrão. A base de dados utilizada é composta por 1000 vídeos, igualmente divididos entre Legítimos e *Stuffing*. Foi adotada a estratégia de validação cruzada, utilizando 05 *folds* em sua execução, cada um com 800 vídeos para treinamento e 200 para testes, igualmente distribuídos entre as duas categorias.

A Figura 2 apresenta o gráfico com os resultados obtidos para cada um dos descritores avaliados. Dada a natureza dos vídeos *stuffing*, pouca variabilidade visual entre os quadros e curta duração, dentre outras, o STIP, algoritmo que considera o vídeo como um todo em sua tarefa de descrição de conteúdo, obteve melhor desempenho, com base nos dois parâmetros verificados: Taxa de Verdadeiros Positivos (TPR, do inglês *True Positive Rate*) e Taxa de Falsos Positivos (FPR, do inglês *False Positive Rate*). O valor obtido de 98% para a identificação de Verdadeiros Positivos, indica grande poder de detecção de vídeos do tipo *stuffing*. Contudo, o índice de 12,6% de Falsos Positivos, ainda representa um grande obstáculo para adoção da metodologia em mecanismos automatizados, uma vez que desprezaria um grande número de vídeos legítimos, por classificá-los como não-colaborativo.

### **3.2. Análise comparativa do desempenho da configuração do *kernel* do SVM**

Nesta seção são apresentados os resultados obtidos nos experimentos exploratórios de configuração do *kernel* do classificador SVM aplicados em uma mesma tarefa de classificação. Os experimentos foram realizados utilizando o ambiente Weka e o STIP como descritor de características visuais dos vídeos. A base de dados utilizada é composta por 1000 vídeos, igualmente divididos entre Legítimos e *Stuffing*. Foi adotada a estratégia de validação cruzada, utilizando 05 *folds* em sua execução, cada um com 800 vídeos para treinamento e 200 para testes, igualmente distribuídos entre as duas categorias.

Para avaliar o desempenho de cada configuração, partiu-se de um cenário referência, onde foram analisados os resultados do classificador SVM, utilizando cada um de seus diferentes *kernels* considerados neste trabalho: Linear, RBF e Polinomial, todos com suas configurações padrão, conforme implementação da biblioteca LibSVM (Chang e Lin, 2011). Os resultados obtidos no cenário de configurações padrão são exibidos na Figura 3. Com base nos parâmetros de TPR e FPR, o melhor desempenho verificado foi o do *kernel* Linear, 98% (TPR) e 12,6% (FPR).

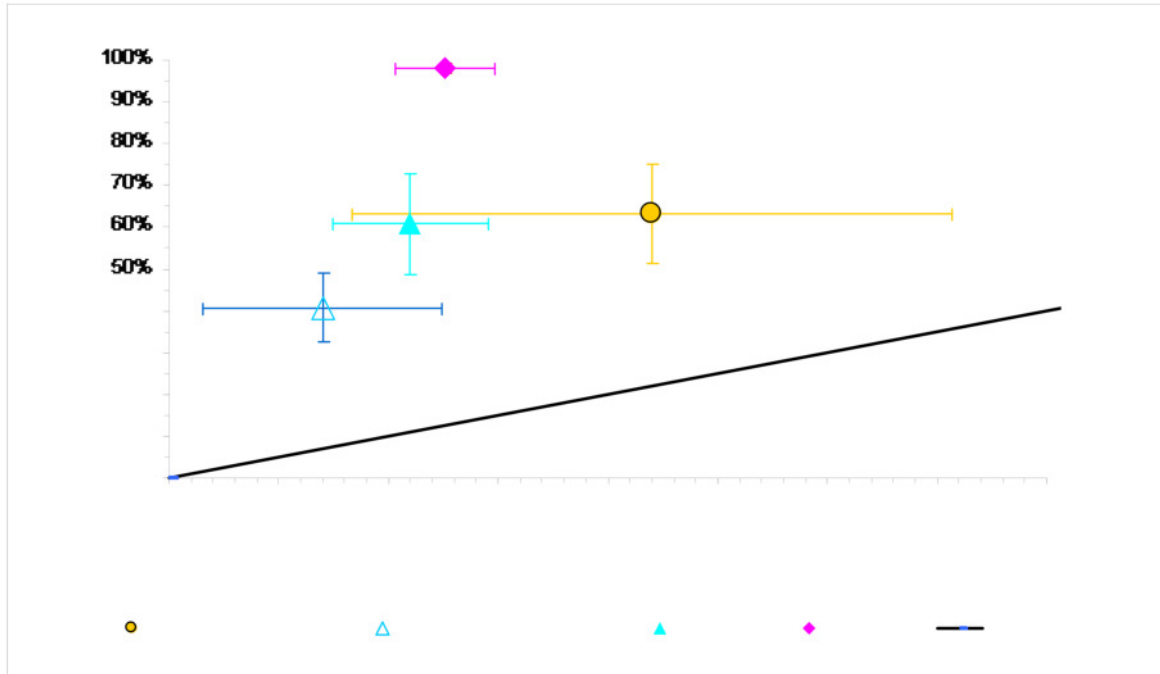


Figura 2. Resultados de TPR e FPR na identificação de vídeos *Stuffing* x Legítimos, utilizando os algoritmos Histograma de Cor nos canais RGB, Momentos de Zernike, SIFT e STIP; e SVM com *kernel* linear.

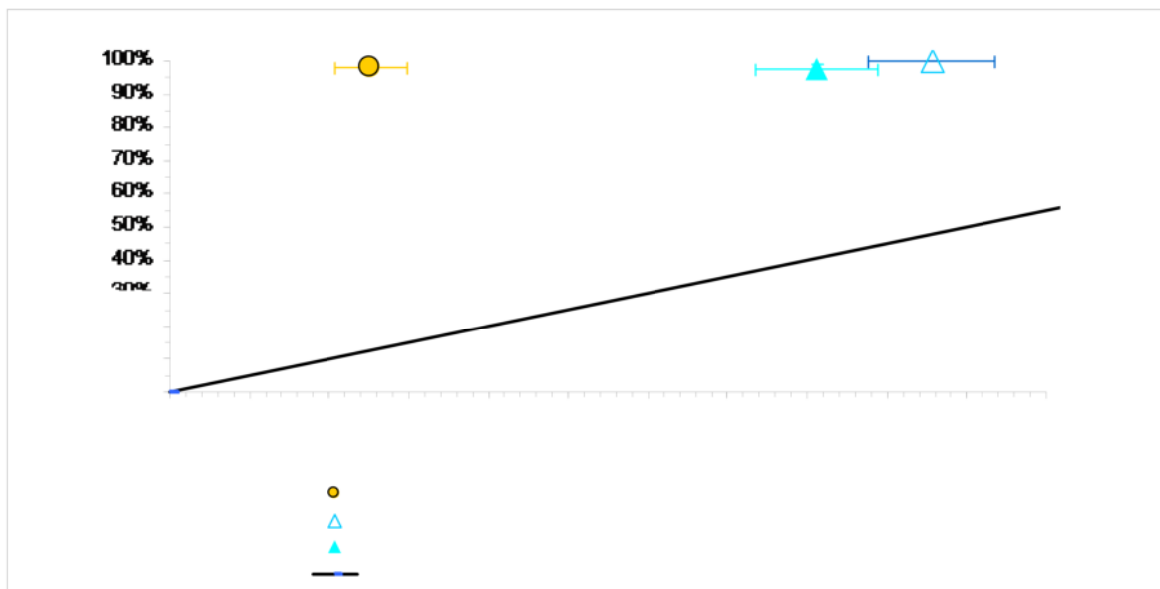


Figura 3. Resultados de TPR e FPR na identificação de vídeos *Stuffing* x Legítimos, utilizando o algoritmo STIP e classificador SVM com diferentes funções *kernel*: Linear, RBF e Polinomial. Todas com suas configurações padrão, conforme implementação da biblioteca LibSVM.

Dando prosseguimento com as investigações, foram realizados experimentos com o *kernel* RBF, manipulando o parâmetro *gamma*. O referido parâmetro possui um valor positivo, entre 0 e 1, sendo por padrão definido com o valor  $1/n^\circ$  de dimensões. Para o experimento construído, este

valor padrão se aproxima de 0,0002. Partindo desta informação, foram investigados valores distintos de  $\gamma$ , variando de 0,2 à 0,0002.

A Figura 4 apresenta os resultados obtidos. Os experimentos com  $\gamma = 0,02$  e  $\gamma = 0,002$ , obtiveram melhores resultados.  $\gamma = 0,02$ , 95,4% (TPR) e 18,2 (FPR).  $\gamma = 0,002$ , 100% (TPR) e 30,8% (FPR).

Com base na análise dos resultados obtidos com a manipulação do parâmetro  $\gamma$  no *kernel* RBF, foram executados experimentos utilizando o *kernel* Polinomial, que recebe a influência de dois parâmetros principais:  $\gamma$  e *degree* (grau do polinômio). O valor padrão para o parâmetro de grau do polinômio é 3. Mantendo o valor de  $\gamma$  constante, o valor de grau de polinômio foi variado de 2 à 4. As Figuras 5 e 6 apresentam os resultados obtidos para os experimentos em que foram variados os valores do grau do polinômio e mantidos os valores de  $\gamma$  constantes em 0.002 e 0.02, respectivamente.

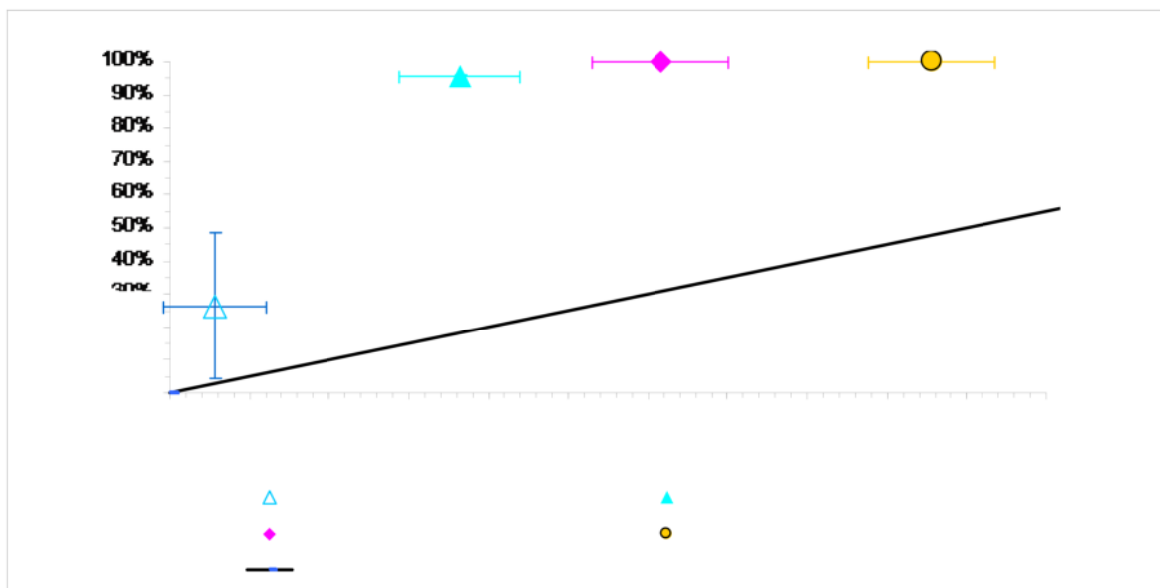


Figura 4. Resultados de TPR e FPR na identificação de vídeos *Stuffing* x Legítimos, utilizando o algoritmo STIP e classificador SVM com diferentes configurações para o parâmetro  $\gamma$  na função *kernel* RBF.

O melhor desempenho com  $\gamma$  igual a 0.002 foi obtido com o *degree* igual a 4, sendo 94,6% (TPR) e 8,8% (FPR). Já com  $\gamma$  igual a 0.02, o melhor desempenho foi obtido com *degree* igual a 2, 95,6% (TPR) e 8,8 (FPR).

A identificação de um melhor resultado é subjetiva, dependente da natureza e objetivos da tarefa de classificação, pois deve levar em consideração o compromisso entre a Taxa de Verdadeiros Positivos e a Taxa de Falsos Positivos. Contudo, os resultados encontrados

demonstram ganho significativo de qualidade em relação a configuração padrão do *kernel* Polinomial.

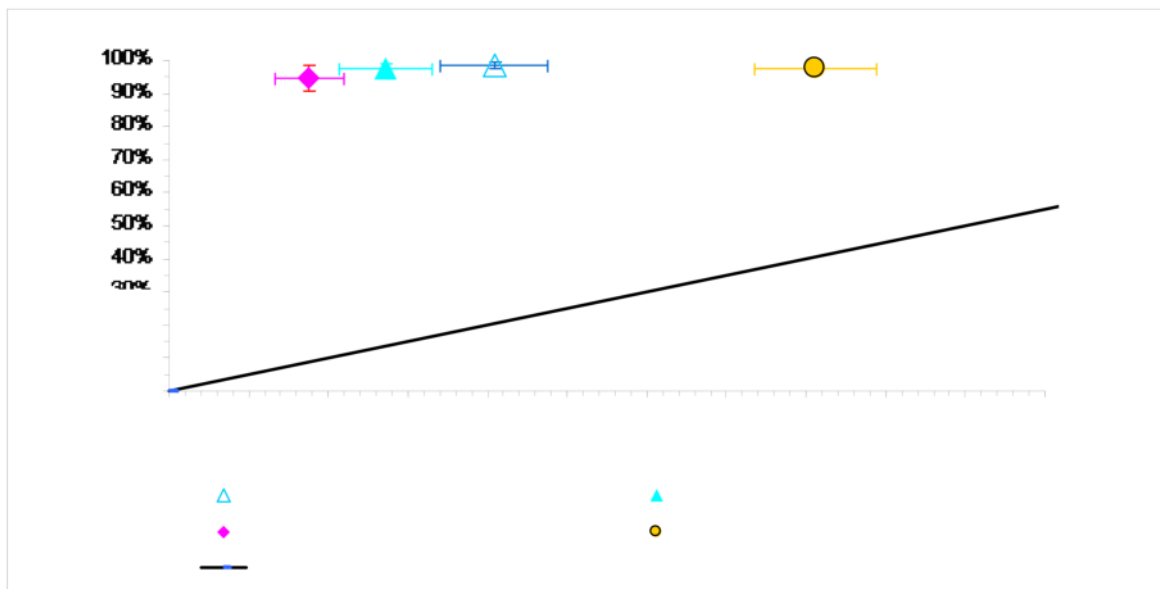


Figura 5. Resultados de TPR e FPR na identificação de vídeos *Stuffing* x Legítimos, utilizando o algoritmo STIP e classificador SVM com diferentes configurações para o parâmetro *degree* na função *kernel* Polinomial. O parâmetro *gamma* foi fixado em 0.002.

## 5 CONSIDERAÇÕES FINAIS

O presente projeto teve como objetivo principal a formação de pessoal qualificado na área de Processamento de Imagens e Análise Semântica de Conteúdo Visual. Considerando essa perspectiva, durante a realização de todas as etapas do projeto primou-se pela reflexão acerca dos cenários encontrados, de modo que tivessem conhecimento das motivações das decisões.

Foram vivenciadas dificuldades de infraestrutura, concernente ao armazenamento e processamento de informação, além de prejuízos advindos da inconstância do serviço de internet. Contudo, os resultados obtidos nos experimentos permitiram avançar o estado da arte para o tipo de problema enfrentado, bem como, possibilitaram aos pesquisadores adquirirem conhecimento prático de como executar um projeto de investigação científica.

## REFERÊNCIAS

Backstrom, L.; Huttenlocher, D.; Kleinberg, J.; Lan, X. **Group formation in large social networks: membership, growth, and evolution.** ACM SIGKDD, pp. 44--54, Philadelphia, PA, USA, 2006.

Chang, C.-C.; Lin, C.-J. **LIBSVM: A library for support vector machines**. ACM Transactions on Intelligent Systems and Technology, vol. 2, 3, 1--27, 2011.

Gonzalez, R. C.; Woods, R. E. **Digital Image Processing**. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 3 edição, 2006.

Hall, Mark; Frank, Eibe; Holmes, Geoffrey; Pfahringer, Bernhard; Reutemann, Peter; Witten, Ian H. **The WEKA Data Mining Software: An Update**. SIGKDD Explorations, Volume 11, Issue 1, 2009.

Khotanzad, A.; Hong, Y. H. **Invariant image recognition by zernike moments**. IEEE Trans. Pattern Anal. Mach. Intell., 12(5):489—497, 1990.

Laptev, I. **On Space-Time Interest Points**. Int. J. Comput. Vision 64, 2-3 (Sep. 2005), 107-123, 2005.

Lowe, D. G. **Object recognition from local scale-invariant features**. IEEE International Conference on Computer Vision, vol. 2, p. 1150, Washington, DC, USA, 1999.

Luz Jr., Antonio. **Deteção de vídeos não-colaborativos com base no conteúdo visual em redes sociais para compartilhamento de vídeo**. Universidade Federal de Minas Gerais (*tese*), 2012.

Newman, M. E. J.; Park, J. **Why social networks are different from other types of networks**. Journal of Phys. Rev., vol. 68, 3, 2003.

Vapnik, V. **Estimation of Dependences Based on Empirical Data**. Nauka, Moscow. [*in Russian*] (English translation: 1982, Springer Verlag, New York), 1979.

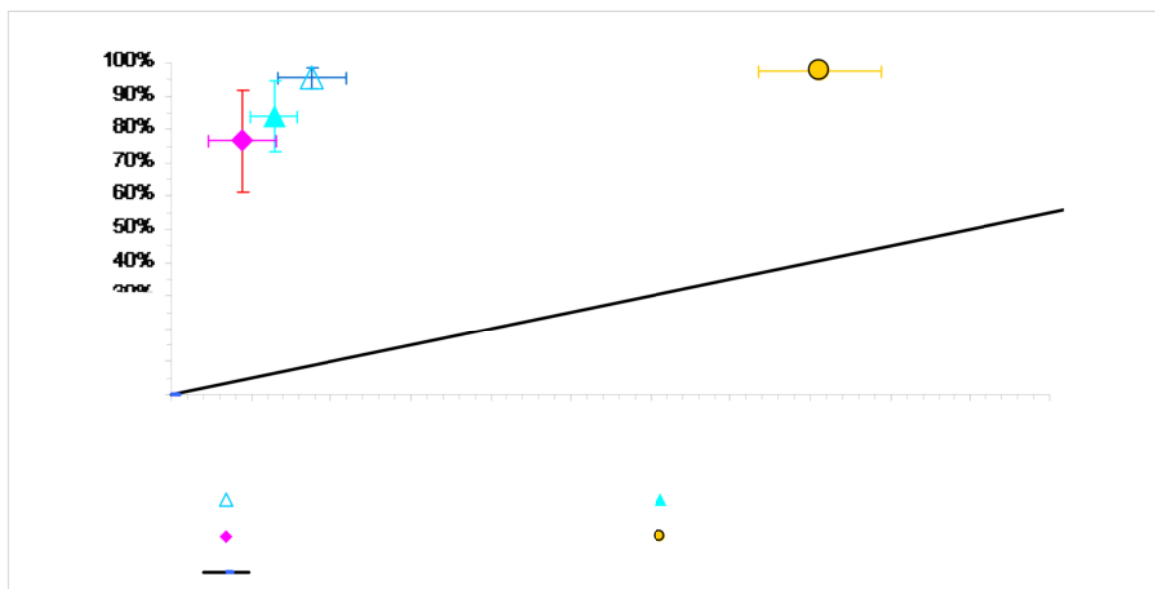


Figura 6. Resultados de TPR e FPR na identificação de vídeos *Stuffing* x Legítimos, utilizando o algoritmo STIP e classificador SVM com diferentes configurações para o parâmetro *degree* na função *kernel* Polinomial. O parâmetro *gamma* foi fixado em 0.02.