

RECONHECIMENTO AUTOMÁTICO DE TATUAGENS: APLICAÇÃO NA ÁREA FORENSE

AUTOMATIC TATTOO RECOGNITION: APPLICATION IN THE FORENSIC FIELD

RECONOCIMIENTO AUTOMÁTICO DE TATUAJES: APLICACIÓN EN EL ÁMBITO FORENSE



10.56238/revgeov16n5-044

Marcelo Musci

Doutor em Engenharia Elétrica

Instituição: Universidade do Estado do Rio de Janeiro (UERJ)

E-mail: marcelo.musci@uerj.br

Flavio Lucas dos Santos Baptista

Graduado em Análise e Desenvolvimento de Sistemas

Instituição: Centro Universitário ETEP (UERJ)

E-mail: flavioagamotto@gmail.com

Carlos Augusto Sicsú Ayres do Nascimento

Doutor em Computação de Alto Desempenho - Multidisciplinar (COPPE/UFRJ)

Instituição: Universidade Federal do Rio de Janeiro (UFRJ)

E-mail: carlos.nascimento@uerj.br

Giancarlo Cordeiro da Costa

Doutor em Engenharia Mecânica Computacional

Instituição: Universidade Federal do Rio de Janeiro (UFRJ)

E-mail: giancarlo.costa@uerj.br

RESUMO

Este trabalho apresenta um sistema de reconhecimento de imagens de tatuagens de forma automática em um banco de dados, que pode vir a ser utilizado para ajudar na correta identificação de criminosos que possuam esse tipo de marca corporal, seu significado e a possível identificação de facções criminosas aos quais pertencem. As técnicas de extração de parâmetros e reconhecimento de imagens são baseadas nos algoritmos Scale Invariant Feature Transform (SIFT), Speed-Up Robust Feature (SURF) e rede neural YOLO. No trabalho são descritos os casos de teste que devem ser seguidos e que irão permitir a comparação de desempenho consistente de métodos de reconhecimento de tatuagens. Os resultados mostraram uma boa performance do algoritmo SIFT (88,65%) em comparação com os demais métodos analisados.

Palavras-chave: SIFT. YOLO. SURF. Tatuagens. Reconhecimento.



ABSTRACT

This work presents a system for the automatic recognition of tattoo images within a database, which can be used to assist in the accurate identification of criminals who bear such body marks, their meanings, and the possible identification of criminal factions to which they may belong. The techniques for parameter extraction and image recognition are based on the Scale Invariant Feature Transform (SIFT) algorithm, the Speeded-Up Robust Features (SURF) algorithm, and the YOLO neural network. The study describes the test cases that must be followed, which will allow for consistent performance comparison of tattoo recognition methods. The results showed good performance of the SIFT algorithm (88.65%) compared to the other methods analyzed.

Keywords: SIFT. YOLO. SURF. Tattoo. Recognition.

RESUMEN

Este trabajo presenta un sistema de reconocimiento de imágenes de tatuajes de forma automática en una base de datos, que puede ser utilizado para ayudar en la correcta identificación de criminales que posean este tipo de marca corporal, su significado y la posible identificación de las facciones delictivas a las que pertenecen. Las técnicas de extracción de parámetros y reconocimiento de imágenes se basan en los algoritmos Scale Invariant Feature Transform (SIFT), Speed-Up Robust Feature (SURF) y la red neuronal YOLO. En el trabajo se describen los casos de prueba que deben seguirse y que permitirán la comparación de desempeño consistente de los métodos de reconocimiento de tatuajes. Los resultados mostraron un buen rendimiento del algoritmo SIFT (88,65%) en comparación con los demás métodos analizados.

Palabras clave: SIFT. YOLO. SURF. Tatuagens. Reconhecimento.



1 INTRODUÇÃO

Tatuagens podem ser uma ferramenta muito útil

1 INTRODUÇÃO

Tatuagens podem ser uma ferramenta muito útil para identificação de pessoas em aplicações criminais, sendo inclusive utilizadas para identificação de vítimas ou suspeitos usando identidades falsas. Seu uso tem sido mais difundido na atualidade, quando outros tipos de identificação biológicos, como impressões digitais, não estão disponíveis.

Para uso em identificação criminal, seu uso possibilita a obtenção de informações discriminatórias como grupo criminoso, religião, antecedentes criminais e outras.

Embora uma tatuagem por si só possa não ser suficiente para diferenciar entre dois indivíduos, pode ajudar a estreitar a identidade de suspeitos ou vítimas nas investigações. Uma tatuagem é uma opção e um traço biométrico que pode conter mais características discriminativas e recursos para identificação de pessoa do que a biometria suave tradicional, como idade, sexo, raça, altura e peso (Beauthier; Lefevre; Valck, 2011).

Além disso, tatuagens podem fornecer informações valiosas sobre as afiliações de um indivíduo ou crenças expressas e pode apoiar a verificação da identidade de um indivíduo.

Em desenvolvimento desde 2014, em um trabalho conjunto com o Instituto Nacional de Padrões e Tecnologia (NIST), o FBI utiliza uma tecnologia, chamada apenas de “Tatt-C”, capaz não apenas de reconhecer tatuagens, mas também de identificar seu significado e com isso descobrir inclinações pessoais, políticas, sociais e até mesmo religiosas de um suspeito.

De acordo com Pocevičė *et al.* (2024), as tatuagens são valiosas para ajudar a identificar associações, grupos, membros, gangues, criminosos ou vítimas. As tatuagens são consideradas biometria suave porque, ao longo do tempo, podem sofrer alterações no corpo humano, diferentemente de características biométricas rígidas como impressões digitais ou íris (Silva; Lopes, 2021). No entanto, a precisão da identificação e detecção automática de tatuagens é desafiada por uma ampla variedade de composições artísticas, cores, formas, texturas, condições e qualidade da imagem (Sun *et al.*, 2016).

Embora algumas pesquisas tenham sido feitas na área de reconhecimento de tatuagens com base em imagens (Pocevičė *et al.*, 2024; Jain, Lee, Jin, 2007; Acton, Rossi, 2008; J. Allen *et al.*, 2011; ANSI/NIST-ITL 1-2011, 2015; Manger, 2012; Heflin, Scheirer, Boulton, 2012; Duangphasuk, Kurutach, 2013; Di; Patel, 2016), esse tipo de estudo ainda não é um produto consolidado na indústria e merece um estudo mais detalhado para o desenvolvimento de sistemas para aplicações governamentais de



próxima geração.

As tatuagens são valiosas para ajudar a identificar associações, grupos, membros, gangues, criminosos ou vítimas. As tatuagens são consideradas biometria suave porque, ao longo do tempo, podem sofrer alterações no corpo humano, diferentemente de características biométricas rígidas como impressões digitais ou íris (Silva; Lopes, 2021). No entanto, a precisão da identificação e detecção automática de tatuagens é desafiada por uma ampla variedade de composições artísticas, cores, formas, texturas, condições e qualidade da imagem (Sun *et al.*, 2016).

A natureza operacional do problema proposto neste trabalho impõe desafios sobre as metodologias tradicionais de recuperação de imagens, dada a grande variação no ambiente e no processo de captura e no conteúdo e qualidade das tatuagens. São exemplos de tais desafios:

- Variação da iluminação e escala da imagem;
- Oclusões parciais de roupas e imagens de fundo com ruídos;
- Variações ao redor das tatuagens;
- Imagens mescladas ou várias tatuagens em uma única imagem;
- Orientação inconsistente da imagem (rotação e cisalhamento);
- Tatuagens extremamente desbotadas;
- Abstrações ambíguas ou desconhecidas.

Técnicas de reconhecimento de tatuagens, para uso de identificação criminal, são baseadas em fotografias e descrições resumidas, como o descrito em (ANSI/NIST-ITL 1-2011, 2016). Nesse sistema uma busca por tatuagens é feita de forma manual, através de palavras chaves inserida no contexto de cada tatuagem. Esse sistema tende a ser bastante rudimentar e lento na busca e identificação de tatuagens, pois depende de uma busca manual exaustiva. Sistemas de reconhecimento automático podem ser implementados de forma a facilitar esse trabalho.

No trabalho de Sun *et al.* (2016) foi proposto um sistema de reconhecimento de imagens, baseado na extração de características visuais como cor textura e forma para identificação de imagens.

Como atualmente o número de pessoas que utilizam tatuagens é bastante significativo, principalmente entre criminosos, um sistema para determinar, tatuagens, de forma rápida e eficiente, se torna bastante promissor para a sociedade.

A ideia é que o sistema proposto possa identificar em cenas criminais, através de imagens de câmeras de segurança, possíveis tatuagens em pessoas que possam levar a sua identificação. Além de identificação de inclinações religiosas, políticas, pessoais e sociais, em tatuagens de criminosos, o que significa que o método desenvolvido não servirá como um simples banco de dados.



2 REFERENCIAL TEÓRICO

2.1 ALGORITMO SIFT

As características das imagens foram obtidas com a utilização do algoritmo Scale Invariant Feature Transform (SIFT) que extrai pontos-chave, denominados keypoints, da imagem em análise e gera descritores sobre esses pontos que descrevem a textura ao seu redor. Esses pontos-chave são invariantes a escala e rotação da imagem e suficientemente robustos quanto a distorções, ângulo de visão e ruídos. O trabalho proposto em (Jain; Lee; Jin, 2007) utiliza o mesmo algoritmo para extração de características, obtendo excelentes resultados.

O algoritmo SIFT corresponde a uma técnica que permite extrair as características distintivas de imagens, sendo bastante usado para o reconhecimento e classificação de objetos em diferentes pontos de vista. Neste caso, verificamos que as características extraídas são invariantes em relação à escala e orientação da imagem.

O processo de extração de características passa por quatro fases. A primeira corresponde à fase onde são detectadas as localizações de possíveis pontos de interesse invariantes à escala e orientação. Consistindo em se detectar extremos (máximos e mínimos) em uma pirâmide da imagem convoluída com a função Diferença de Gaussiana (DoG - Difference of Gaussian). Pontos chave correspondem a estes extremos para diferentes escalas.

Um filtro Gaussiano passa baixa é dado pela convolução de uma imagem I com a função G :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

onde:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

A função DoG é dada pela diferença de imagens filtradas em escalas próximas separadas por uma constante k . O resultado de se fazer a convolução de uma imagem com o filtro DoG é dado por:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3)$$

ou seja, é a diferença entre imagens borradas por um filtro gaussiano em escalas σ e $k\sigma$. Este filtro consegue detectar variações de intensidade na imagem, tais como contornos. Variando-se o σ , é possível encontrar descritores para variações em diferentes escalas espaciais. Lowe (2004) considera que é necessário fazer a convolução da imagem até 2σ para ser possível a construção de descritores invariáveis à escala. Ao conjunto DoG e as imagens suavizadas é dado o nome de oitava.



Nesse trabalho foram utilizados os parâmetros padrão descritos por Lowe na utilização do SIFT, como número de oitavas e intervalos.

Na segunda fase, os pontos candidatos detectados anteriormente são examinados com maior detalhe e filtrados. Neste caso, é analisada com mais detalhe a informação dos vizinhos e um modelo é ajustado de modo a determinar corretamente a localização e escala. Com esta informação são descartados os pontos com baixo contraste e que desta forma são sensíveis ao ruído.

Na fase seguinte é definida uma orientação a cada ponto-chave. Neste caso, todas as propriedades do ponto-chave são medidas relativamente à orientação deste, possibilitando, assim, que estas sejam invariantes em relação à rotação.

O último passo consiste no cálculo de um descritor associado a cada um dos pontos chave. Este descritor corresponde a um vetor de tamanho fixo que é normalizado de forma a ser invariante em relação a mudanças de iluminação, a Figura 1 mostra uma imagem e seus pontos-chaves extraídos pela utilização do SIFT.

Assim, com base nos descritores extraídos, é possível realizar a correspondência entre duas imagens.

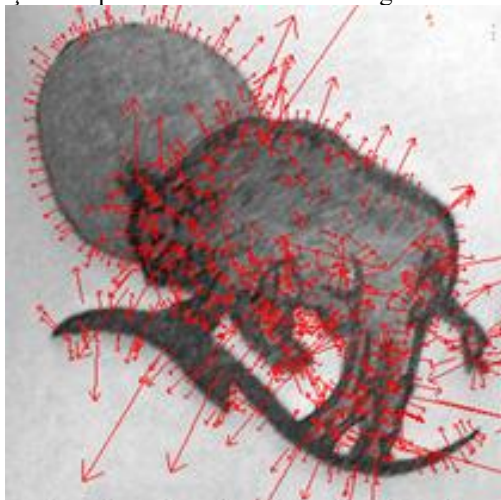
O processo de comparação de duas imagens torna-se uma comparação de pontos-chave baseada em seus descritores.

O melhor candidato para cada ponto-chave é encontrado através da identificação de seu vizinho mais próximo na imagem de referência, armazenada no banco de dados, que contém todos os pontos-chave de cada imagem a ser comparada. Os vizinhos mais próximos são definidos como os pontos-chave, com distância mínima euclidiana do vetor descritor dado. A probabilidade de que uma correspondência está correta pode ser determinada utilizando a relação de distância do vizinho mais próximo com a distância do segundo vizinho mais próximo.

No sistema proposto todas as correspondências em que a relação distância for superior a 0,8 serão rejeitadas, o que segundo Lowe (2004) elimina 90% das falsas correspondências enquanto descartando menos de 5% dos resultados corretos.



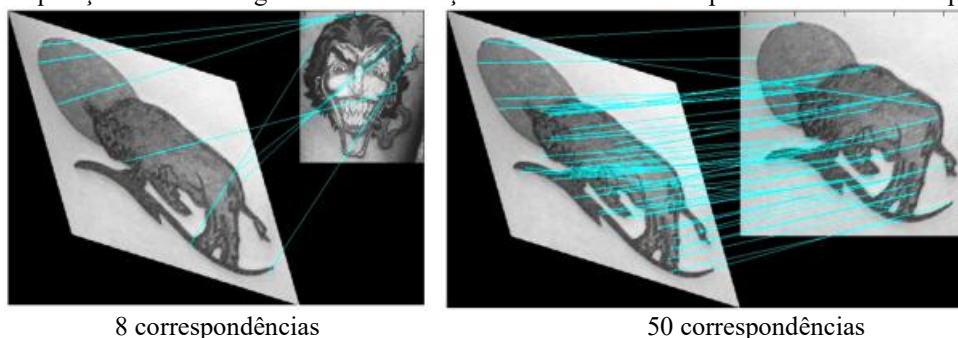
Figura 1. Localização dos pontos-chave em uma imagem com a utilização do SIFT.



Fonte: Os autores, (2021).

A métrica utilizada para comparação entre as imagens é baseada no número de correspondências entre os pontos-chave das duas imagens em análise, conforme visto na Figura 2.

Figura 2. comparação de duas imagens com a utilização do número de correspondências entre seus pontos-chave.



8 correspondências

50 correspondências

Fonte: Os autores, (2021).

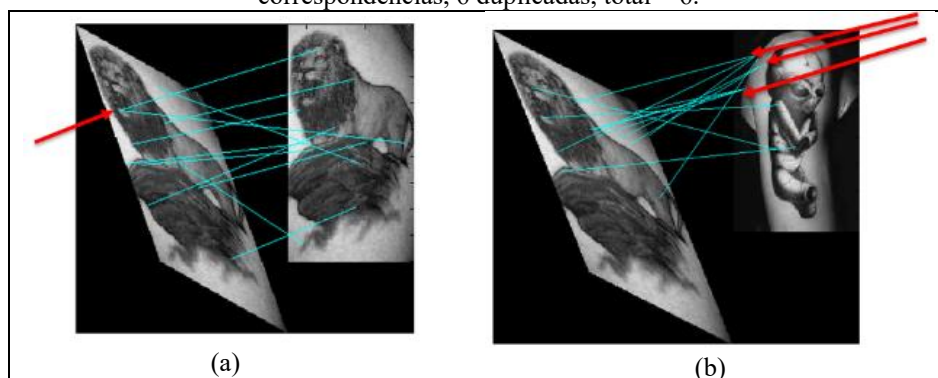
Na comparação entre duas imagens podem ocorrer múltiplas correspondências entre pontos-chave distintos, que possivelmente levam a erros na busca correta da imagem, devido a geração de um número elevado de falsas correspondências em imagens distintas, isso pode ser especialmente notado em imagens muito distorcidas e com poucos pontos-chave. Neste trabalho o algoritmo original SIFT foi modificado de forma a se eliminar essas múltiplas correspondências, aumentando assim a eficácia do sistema proposto.

Na comparação entre duas imagens podem ocorrer múltiplas correspondências entre pontos-chave distintos, que possivelmente levam a erros na busca correta da imagem, devido a geração de um número elevado de falsas correspondências em imagens distintas, isso pode ser especialmente notado em imagens muito distorcidas e com poucos pontos-chave.

Ao se eliminar essas múltiplas correspondências pode-se aprimorar a escolha correta da imagem.

Conforme pode ser visto na Figura 3(a), que apresenta uma imagem onde ocorrem onze correspondências, sendo uma duplicada e na imagem apresentada na Figura 3(b) ocorrem doze correspondências e seis duplicadas, a imagem a ser escolhida como certa seria a 3(b) - maior número de correspondências, porém com a eliminação das correspondências duplicadas consegue-se selecionar a imagem correta.

Figura 3 – Eliminação das correspondências múltiplas. Em (a) 11 correspondências, 1 duplicada, total = 10, em (b) 12 correspondências, 6 duplicadas, total = 6.



Fonte: Os autores, (2021).

2.2 ALGORITMO SURF

Proposto por Bay, Tuytelaars e Van Gool (2006), o descritor Características Robustas Aceleradas (Speed-Up Robust Feature - SURF) apresenta semelhanças com relação ao SIFT e também permite detecção e descrição de regiões de interesse. Os autores afirmam que SIFT apresentou o melhor desempenho dentre os descritores analisados. Dentre suas vantagens, estão a capacidade de representar informações dos padrões espaciais e robustez a pequenas deformações e erros de localização das regiões detectadas. Porém, SIFT apresenta alta dimensionalidade em seu descritor e seu custo computacional dificulta seu uso para aplicações on-line.

Seguindo esta linha de raciocínio, no trabalho descrito em (Bay; Tuytelaars; Van Gool, 2006), foi proposto um descritor com a metade do número de elementos do SIFT e baseado em matrizes Hessianas, as quais garantem maior estabilidade que detectores de cantos de Harris, além de apresentar boa performance em termos de tempo computacional e taxas de acerto. A proposta ainda inclui o uso de determinantes da matriz Hessiana, devido a sua robustez quanto a estruturas mal localizadas e alongadas. Basicamente, as estruturas detectadas pelo SURF estão situadas em regiões que maximizam o determinante da matriz Hessiana.

O uso do conceito de imagens integrais reduz drasticamente o tempo de computação devido aos filtros de convolução baseados em caixas. Por definição, cada elemento de uma imagem integral I (x,y) representa a soma de todos os pixels da imagem original I contidos na região retangular situada entre a origem e (x, y), conforme definição na equação a seguir. Diante disto, o cálculo da soma dos pixels contidos em uma região qualquer requer apenas três operações de adição, o tempo necessário



para aplicar qualquer filtro à imagem permanece constante (independente de suas dimensões), todos os filtros são aplicados na imagem integral sem a necessidade da criação de novas imagens por meio de interpolações, os componentes de alta frequência são preservados e poderão ser recuperados independentemente da escala corrente (Bay; Tuytelaars; Van Gool, 2006).

Semelhante à informação do gradiente extraída pelo SIFT, o descritor SURF representa a distribuição dos níveis de cinza na vizinhança do ponto de interesse (x, y). Para isso, SURF considera a derivada de primeira ordem da wavelet de Haar. Uma vez detectada uma região de interesse, sua orientação é determinada pelas respostas da wavelet de Haar nas direções 'x' e 'y' para cada ponto contido na vizinhança do ponto de interesse. Tais respostas, após serem multiplicadas por pesos definidos pela função Gaussiana centrada no ponto de interesse, são representadas como pontos no espaço e acumuladas considerando as direções horizontal e vertical e uma janela deslizante de orientação.

Por fim, o maior valor (vetor), considerando todas as possibilidades geradas durante o deslocamento da janela deslizante de orientação, determina a orientação do ponto de interesse.

Após identificar a região de interesse e sua orientação, o próximo passo consiste em calcular os valores do descritor. Para isso, define-se uma região quadrada (8×8) centrada no ponto de interesse e alinhada com a orientação previamente identificada. Para cada um dos 64 elementos, obtém-se a resposta da wavelet de Haar direções 'x' e 'y', denominadas respectivamente dx e dy . Para cada sub-região 2×2 , dx , dy , $|dx|$ e $|dy|$ são acumuladas separadamente, gerando 16 conjuntos com as quatro características (dx , dy , $|dx|$ e $|dy|$), num total de 64 elementos. Ao final, igual ao SIFT, geralmente são detectados centenas ou até milhares de pontos para a imagem. Todos estes pontos são invariantes à translação, escala e rotação, além de ser parcialmente invariante a mudanças de iluminação e projeções 3D e potencialmente candidatos a compor o conjunto de características que descreve a imagem (Bay; Tuytelaars; Van Gool, 2006).

2.3 REDE YOLO

A rede YOLO divide a imagem em regiões menores e prevê caixas delimitadoras, probabilidades de existir um objeto naquela caixa e também qual a probabilidade de ser determinado objeto (Redmon; Farhadi, 2017).

O YOLO utiliza uma rede neural profunda (DNN – Deep Neural Network), uma rede neural convolucional, com o mesmo nome do framework utilizado para implantá-lo. Sua implementação foi desenvolvida na linguagem C, porém, com a ajuda da comunidade e empresas, já é disponível em várias outras linguagens de programação.

O YOLO é surpreendentemente simples. Uma única rede convolucional prevê simultaneamente múltiplas caixas delimitadoras (bounding boxes) e as probabilidades de classes para essas caixas. O



YOLO treina com imagens completas e otimiza diretamente o desempenho da detecção. Esse modelo unificado apresenta várias vantagens em relação aos métodos tradicionais de detecção de objetos.

Durante o processo de detecção de objetos, podem ser encontradas diversas caixas delimitadoras para um mesmo objeto. Estas caixas também podem possuir diversos tamanhos diferentes. Cada caixa possui uma pontuação de confiança, indicando a quantidade de certeza que o algoritmo tem que aquela caixa possui um objeto.

A versão 12 da YOLO foi lançada em abril de 2024 e suas principais características que podem ser destacadas nessa versão são a melhoria na velocidade de inferência e acurácia. Outra característica importante é o fato de ser mais eficiente, pois foi otimizada para utilizar menos memória. O YOLOv12 demonstrou ser o melhor detector de objetos para testes em tempo real.

O processo de treinamento dos algoritmos pode envolver várias etapas, dependendo de sua característica. Nos algoritmos supervisionados, o algoritmo é treinado usando um grande conjunto de imagens rotuladas, ou seja, imagens previamente classificadas como áreas limpas ou áreas com a presença de resíduos. Essas imagens de treinamento fornecem informações supervisionadas para o algoritmo aprender a reconhecer padrões específicos, o mesmo procedimento foi utilizado em (Musci, *et al.* 2024a , 2024b).

Ao final do treinamento, assim como mencionado em (Di; Patel, 2016), a rede YOLO deve ser capaz de identificar as tatuagens em um banco de dados.

Durante o treinamento de uma rede neural, o valor esperado para cada objeto a ser detectado, em uma iteração, é comparado com os valores de saída. Cada ciclo de treinamento recebe o nome de epoch e, quanto maior o seu número, mais fácil será para a rede trabalhar com os valores de treinamento. É importante lembrar que tanto pouco treino quanto treino em excesso podem ter impacto negativo na rede. Com pouco treino, ela não consegue aprender a lidar com os valores de treino, um problema chamado de underfitting. Com treino em excesso, a rede pode ficar condicionada aos valores de treino, e não conseguirá obter sucesso quando receber valores diferentes, o que é chamado de overfitting. Ainda em seu treinamento, o valor chamado de loss representa a diferença entre o valor de saída da rede e o esperado para uma determinada entrada. A diminuição gradativa do valor de loss a cada epoch indica que a rede está obtendo mais sucessos ao fazer previsões (Filipakis; Fagundes, 2021).

3 METODOLOGIA

Este trabalho descreve testes de comparação de desempenho no método proposto para reconhecimento de tatuagens para uso forense.

A caracterização de tatuagens é quase única, pois mesmo que duas pessoas tenham a mesma tatuagem, a unicidade de cada uma é praticamente garantida, pois são feitas a mão.



O sistema deverá ser robusto o suficiente para desprezar pequenas variações na imagem pesquisada, fornecendo em sua resposta uma lista de tatuagens similares, em ordem decedente, encontradas no seu banco de dados, veja Figura 4, sendo a decisão final, quanto à similaridade, efetuada pelo elemento humano (Musci *et al.*, 2025).



Fonte: Os autores (2021).

A técnica de extração de parâmetros e reconhecimento de imagens é baseada, nos algoritmos Scale Invariant Feature Transform – SIFT e Speeded-UP Robust Features – SURF, que são detectores de características amplamente utilizados na área de correspondência ou reconhecimento de imagens. Esses métodos procuram por pontos significativos (geralmente cantos) na imagem de entrada e são invariantes a transformações geométricas, condições de iluminação ou ruído na imagem. Cada ponto característico é definido de forma única por seu vetor de características de alta dimensão, que descreve sua vizinhança.

O SIFT geralmente extrai mais características do que o SURF, o que lhe confere maior precisão, porém com o custo da velocidade. Em comparação, o SURF é mais rápido, mas extrai menos características e, conseqüentemente, tende a apresentar menor precisão do que o SIFT. É igualmente importante considerar as dimensões das imagens. Em imagens pequenas, o SURF pode ter melhor desempenho que o SIFT devido ao seu tempo de processamento mais rápido, enquanto em imagens grandes o SIFT leva vantagem por sua capacidade de extrair um número maior de características. Em conclusão, SIFT e SURF são ambos algoritmos amplamente utilizados para extração de características. Embora possuam funcionalidades semelhantes, cada um apresenta certas vantagens e desvantagens que tornam um mais adequado que o outro dependendo da aplicação e do tamanho das imagens. Em geral, o SIFT é mais apropriado para imagens grandes e mais complexas, enquanto o SURF é mais indicado para imagens menores e aplicações que exigem processamento mais rápido (Bilik; Horak, 2022; Anandhasilambarasan *et al.*, 2024). De acordo com Juan e Gwun (2009), o SURF é rápido e apresenta bom desempenho, assim como o SIFT, mas não é estável em relação a rotações e variações de iluminação.

Outra ferramenta de Visão Computacional que tem ganhado muita atenção nos últimos anos é uma rede neural convolucional (CNN) denominada YOLO: You Only Look Once (Redmon; Farhadi, 2017).

Após o seu lançamento em 2015, o YOLO foi logo reconhecido como uma técnica inovadora pois através de uma abordagem totalmente nova foi capaz de obter uma precisão igual ou superior aos outros métodos de detecção de objetos da época, porém com uma velocidade de detecção muito superior. Os sistemas de detecção de objetos anteriores ao YOLO, redirecionam classificadores para realizar a detecção, aplicando um algoritmo a uma imagem em vários locais e escalas, onde as regiões de alta pontuação da imagem são consideradas detecções. Essas regiões menores, extraídas da imagem, são denominadas Bounding Box – “Regiões Delimitadoras” e cada uma dessas regiões detecta um único objeto. Essa técnica de classificação percorre a imagem em várias regiões e escalas diferentes para realizar a classificação do objeto.

3.1 AVALIAÇÃO DOS ALGORITMOS

De forma a se testar o sistema, diversas alterações nas imagens originais podem ser realizadas por intermédio de transformações como: borramento, alteração na luminosidade, rotação, escala e cisalhamento.

Com as imagens do Banco de Dados, descrito anteriormente, recortes manuais foram feitos para isolar as tatuagens e a seguir aplicadas as cinco transformações, gerando assim um total de 9490 imagens para teste do sistema.

As operações realizadas podem ser visualizadas na Figura 5, que mostra diferentes imagens pertencentes a mesma classe.

Na utilização da rede YOLO a métrica utilizada foram a precisão, o recall e a mAP@50.

Figura 5. Operações realizadas nas imagens para teste do sistema proposto.



Fonte: Os autores, (2021).

A precisão indica a proporção de detecções que realmente correspondem ao objeto de interesse. Em outras palavras, mede a qualidade das predições positivas do modelo. Quanto maior a precisão, menor o número de falsos positivos.

$$Precisão = \frac{TP}{TP+FP} \quad (4)$$

O recall mede a capacidade do modelo em identificar corretamente todos os objetos presentes na imagem. Avalia a proporção de objetos reais que foram detectados. Quanto maior o recall, menor o número de falsos negativos.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

O mAP@50 (Mean Average Precision at IoU = 0.5) é a métrica mais utilizada para avaliar modelos de detecção de objetos, pois considera simultaneamente precisão e recall. Ele é calculado a partir da área sob a curva de Precisão-Recall para cada classe e, em seguida, tira-se a média entre todas as classes.

O sufixo @50 significa que uma predição é considerada correta quando a interseção entre a caixa delimitadora prevista e a real (IoU – Intersection over Union) é igual ou superior a 50%.

$$mAP@50 = \text{média das precisões médias (AP) com IoU} \geq 0.5 \quad (6)$$

A máquina utilizada pelos algoritmos é um computador com processador Intel Core i9 14900F, 24 cores de até 5.8GHz, memória Kingston de 128GB/DDR5/5200GHz e GPU RTX 4090/24GB.

4 RESULTADOS E DISCUSSÕES

4.1 BANCO DE DADOS PARA TESTE DO SISTEMA PROPOSTO

O uso de sites de mídia social como meio de dados da coleção foi utilizado para testar o sistema, além disso, as informações coletadas nas redes sociais são frequentemente usadas para identificar um indivíduo. Sendo esse um dos produtos provenientes deste trabalho, ou seja, a criação de um banco de dados disponível para testes, pois não se encontrou base de dados pública para testes. Embora previsto em lei no Brasil pela Lei 12.037/09, a única notícia de criação de uma base de dados encontrada foi do Sistema Phoenix da Secretaria de Segurança Pública de São Paulo (Grosso, 2012).

Assim, foi construído um banco de dados de tatuagens, contendo inicialmente 1898 imagens, com foco em tipos de tatuagens elencados em (Lordelo, 2018), que define as tatuagens utilizadas no mundo do crime, que são:



- Papa-Léguas: distribuição de drogas, geralmente com uso de motos;
- Personagem Taz: significa criminoso acostumado a assalto à mão armada e arrastões;
- Teia de Aranha: criminosos que agem em grupo;
- Águia: significa liberdade e é realizada quando o detento está em cárcere;
- Cruz: criminoso que já foi preso várias vezes;
- Folha de Maconha: usuário de drogas;
- Mulher Nua com genitália de fora: viciado em drogas injetáveis;
- Tigre: caçador implacável com os inimigos;
- Texto “Vida Loka”: criminoso que não mede as consequências dos atos de violência que pratica;
- Cruz com caveira: sinal de lealdade e respeito entre criminosos, mas também pode indicar alguém que matou para não morrer, ou seja, agiu em legítima defesa;
- Imagem do diabo: criminosos pistoleiros e matador que tem fama de ter pacto com o demônio;
- Caveira com faca no crânio: assassinos contumazes, como também matadores de policiais;
- Palhaço e o personagem Coringa: associados à morte de policiais;
- Duendes, Magos e Saci-Pererê: traficantes de drogas;
- Chuck – O Brinquedo Assassino: esse tipo de tatuagem representa criminoso violento com várias passagens por assaltos e homicídios;
- Rosto de Índia: significa matador de policial;
- Desenho da Morte: atrelado a um grupo de extermínio e morte de inimigos;
- Túmulo: aquele que sabe guardar segredo, qualidade valorizada entre os presidiários;
- Pistola ou Revólver: bandido praticante de latrocínio, ou seja, roubo seguido de morte;
- Carpa: ligado ao tráfico de drogas e formação de quadrilha.

4.1.1 Tatuagens Discriminatórias

- Frase “Amor Só de Mãe”: significa longos anos de servidão sexual em cadeias;
- Pênis desenhado nas costas: estuprador;
- Serpente: preso não confiável.

4.1.2 Tatuagens que Indicam Facções Criminosas

- Primeiro Comando da Capital (PCC): Carpa, Escorpião, inscrição “PCC/1533”, Yin Yang, inscrição “Paz Justiça Liberdade”;
- Comando Vermelho (CV): Inscrição “CV” ou “CVRL”;
- Amigos dos Amigos (ADA): Chuck, Brinquedo Assassino.



4.1.3 Tatuagens que retratam homossexuais na cadeia

- Coração com flecha;
- Beija-flôr;
- Flor;
- Coração transpassado por flecha;
- Coração com a inscrição “AMOR DE MÃE”;
- Borboleta tatuada nas costas ou peito;
- Imagem de São Sebastião.

As imagens foram obtidas na internet e em sites de imagens gratuitas como:

- Pixabay: <http://pixabay.com/>
- RGBStock: <http://www.rgbstock.com/>
- Flickr: <http://www.flickr.com/>
- Free Digital Photos: <http://www.freedigitalphotos.net/>
- FreeRange Stock: <http://freerangestock.com/>
- Pexels: <http://www.pexels.com/>
- Unsplash: <http://unsplash.com/>
- Morguefile: <http://morguefile.com/>
- Free Images: <http://pt.freeimages.com/>
- Free Stock Photos Banks: <http://www.freephotosbank.com/>
- Stockvault: <http://www.stockvault.net/>
- Everystockphoto: <http://www.everystockphoto.com/>
- StockSnap.io: <http://stocksnap.io/>
- Public Domain Pictures: <http://www.publicdomainpictures.net/>
- Refe: <http://getrefe.com/>
- Picjumbo: <https://picjumbo.com/>

O banco de dados, um dos produtos provenientes deste trabalho, está disponível no GitLab, em: <https://gitlab.com/tattoo-uerj/reconhecimento-de-tatuagens/-/tree/main>.

4.2 APLICAÇÃO DO ALGORITMO SIFT

No teste das 9490 imagens de tatuagens, foi obtida uma performance de 88,65% em rank-1.

O tempo de execução do sistema, extração das características e análise tatuagens foi de aproximadamente 7 horas.



4.3 APLICAÇÃO DO ALGORITMO SURF

No teste das 9490 imagens de tatuagens, foi obtida uma performance de 64,1% em rank-1.

O tempo de execução do sistema, extração das características e análise tatuagens foi de aproximadamente 2hora e 10 minutos.

4.4 APLICAÇÃO DA REDE NEURAL YOLO

Detectar as tatuagens definidas em (Lordelo, 2021), dentro do bando de tatuagens construído. Foi realizada uma investigação utilizando as classes de tatuagens mostradas na Tabela 1 e suas respectivas taxas de acerto.

Somente foram utilizadas classes com número suficiente de tatuagens para treinamento e teste da rede YOLO.

Tabela 1. Taxa de acerto da rede YOLOv12.

Tipo de tatuagem	Taxa de acerto	Recall	Nota Geral (mAP@50)
papa-léguas	0,67	1,00%	0,995
taz	0,77	0,52	0,79
cruz e cruz com caveira	0,78%	0,25	0,34
aguia	0,64	0,83	0,86
duende	0,74	0,64	0,67
yin yang	0,59	0,25	0,36
índia	0,91	0,89	0,91
morte	0,77	0,60	0,72
tigre	0,83	0,78	0,91
palhaço	0,91	0,47	0,68
chuck	0,69	0,92	0,90
armas	0,69	0,14	0,20
carpas	0,57	0,50	0,59
serpente	0,60	0,53	0,62
borboleta	0,97	0,76	0,87
coração	0,73	0,33	0,35
flor	0,41	0,23	0,32

Fonte: Os autores, (2021).

Um total de 6165 imagens foram usadas para treinamento e 1505 para teste.

O algoritmo de detecção de objetos YOLOv12 foi utilizado na investigação experimental para a detecção de tatuagens. Os modelos pré-treinados do YOLOv12 foram usados como base, disponível em <https://github.com/sunsmarterjie/yolov12>.

O tempo de execução do sistema, para treinamento e análise das 1505 tatuagens, usando o modelo pré-treinado yolov12x.pt, teve duração de 5 horas 17 minutos tendo 217 épocas e um early-stop de 100 sendo a melhor época a 117.

O ambiente utilizado foi o Windows 10, Python 3.11.9, PyTorch 2.8.0+cu126, CUDA 12.6.

Métricas Gerais

Precisão média: 0,658



Recall médio: 0,409

mAP@50 média: 0,53

4.5 DISCUSSÃO DOS RESULTADOS

Os resultados obtidos neste trabalho são os testes utilizando os algoritmos SIFT, SURF e a rede YOLO na detecção dos tipos de tatuagens e a sua possível utilização na identificação forense.

Os valores reportados de desempenho mostram que o sistema baseado no algoritmo de SIFT, com as devidas modificações, apresenta excelentes resultados, 88,65% de acerto das imagens utilizadas, em comparação com o algoritmo SURF, 64,1% e a rede YOLOv12 com 53% de acerto médio. Podendo ser utilizado, com algumas implementações, para a aplicação a que se destina, ou seja, identificação de criminosos por intermédio de suas tatuagens.

Também foi criado um banco de dados com os tipos de tatuagens mais utilizados por criminosos e sua disponibilidade ao público.

5 CONCLUSÃO

O presente estudo teve como objetivo desenvolver um sistema de reconhecimento automático de imagens de tatuagens em banco de dados, empregando os algoritmos SIFT, SURF e a rede neural YOLOv12, visando sua aplicação no processo de identificação de indivíduos envolvidos em atividades criminosas que possuam esse tipo de marca corporal.

Os resultados experimentais evidenciaram a superioridade do algoritmo SURF em relação às demais técnicas analisadas. No entanto, embora apresente maior velocidade de execução, sua eficiência no reconhecimento das imagens avaliadas não se mostrou satisfatória. O uso de redes neurais, a exemplo da YOLOv12, também não demonstrou desempenho superior ao algoritmo SIFT, resultado que pode ser atribuído à insuficiência do conjunto de imagens utilizado para o treinamento.

O tempo relativamente elevado para a identificação das tatuagens pode ser otimizado mediante a implementação do sistema em linguagem compilada, como C, o que reduziria significativamente a latência do processamento.

Em perspectiva futura, o aprimoramento do sistema poderá viabilizar sua integração a câmeras de vigilância urbana ou a drones, ampliando as possibilidades de monitoramento e localização de suspeitos sem a necessidade de confronto armado. Diferentemente de características mutáveis, como cor e corte de cabelo, ou presença de barba e bigode, as tatuagens apresentam maior estabilidade temporal, configurando-se como um recurso confiável para a identificação forense em diferentes contextos.

Assim, o desenvolvimento de um sistema robusto de reconhecimento de tatuagens, capaz de lidar com variações ambientais, representa um avanço significativo para a Ciência Forense no estado



do Rio de Janeiro e em âmbito nacional, oferecendo suporte a ações preventivas e investigativas mais eficazes no enfrentamento da violência urbana.

AGRADECIMENTOS

Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) – PROGRAMA APOIO A PROJETOS DE INOVAÇÃO DASEGURANÇA PÚBLICA - CIÊNCIAS FORENSES 2021 (E_13/2021).



REFERÊNCIAS

- ACTON, S. T.; ROSSI, A. Matching and retrieval of tattoo images: Active contour CBIR and glocal image features. In: 2008 IEEE Southwest Symposium on Image Analysis and Interpretation. IEEE, 2008. p. 21–24. DOI: <https://doi.org/10.1109/SSIAI.2008.4512275>.
- ALLEN, J. D.; ZHAO, N.; YUAN, J.; LIU, X. Unsupervised tattoo segmentation combining bottom-up and top-down cues. In: SPIE 8063: Mobile Multimedia/Image Processing, Security, and Applications. 2011. DOI: <https://doi.org/10.1117/12.884368>.
- ANANDHASILAMBARASAN, D. et al. SIFT and SURF: A Comparative Analysis of Feature Extraction Methods for Image Matching. In: 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2024, Kamand. Proceedings. Kamand: IEEE, 2024. p. 1–6. DOI: <https://doi.org/10.1109/ICCCNT61001.2024.10726049>.
- ANSI/NIST-ITL 1-2011. Update 2016. NIST Special Publication 500-290e3: Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information. National Institute of Standards and Technology, 2015. DOI: <https://doi.org/10.6028/NIST.SP.500-290e3>.
- BAY, H.; TUYTELAARS, T.; VAN GOOL, L. SURF: Speeded up robust features. In: European Conference on Computer Vision – ECCV. 2006. p. 404–417. DOI: https://doi.org/10.1007/11744023_32.
- BEAUTHIER, J. P.; LEFEVRE, P.; VALCK, E. D. Autopsy and identification techniques. In: MÖRNER, N.-A. (ed.). The Tsunami Threat – Research and Technology. London: InTech, 2011. DOI: <https://doi.org/10.5772/13610>.
- BILIK, S.; HORAK, K. SIFT and SURF based feature extraction for the anomaly detection. In: Proceedings I of the 28th Conference STUDENT EEICT 2022. Brno, 2022. p. 459–464. ISBN 978-80-214-6029-4. DOI: <https://doi.org/10.48550/arXiv.2203.13068>.
- DI, X.; PATEL, V. M. Deep Tattoo Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016, Las Vegas. Proceedings. Las Vegas: IEEE, 2016. p. 119–126. DOI: <https://doi.org/10.1109/CVPRW.2016.22>.
- DUANGPHASUK, P.; KURUTACH, W. Tattoo skin detection and segmentation using image negative method. In: 13th International Symposium on Communications and Information Technologies. IEEE, 2013. p. 354–359. DOI: <https://doi.org/10.1109/ISCIT.2013.6645881>.
- FILIPAKIS, G.; FAGUNDES, F. Uma proposta de utilização de Redes Neurais na criação de sequências didáticas baseadas no padrão SCORM. In: XXIII Congresso de Computação e Tecnologias da Informação – Encoinfo. Palmas, 2021. Disponível em: <https://www.researchgate.net/publication/356223762>. Acesso em: 13 mar. 2022.
- GROSSO, E. L. A tecnologia à disposição da Polícia Judiciária. [S.l.]: Conteúdo Jurídico, 2012. Disponível em: <http://www.conteudojuridico.com.br/consulta/Artigos/27992>. Acesso em: 13 mar. 2022.
- HEFLIN, B.; SCHEIRER, W. J.; BOULT, T. E. Detecting and classifying scars, marks, and tattoos found in the wild. In: BTAS '12: IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems. IEEE, 2012. p. 31–38. DOI: <https://doi.org/10.1109/BTAS.2012.6374555>.



- JAIN, A. K.; LEE, J. E.; JIN, R. Tattoo-ID: automatic tattoo image retrieval for suspect and victim identification. In: Pacific-Rim Conference on Multimedia (PCM '07). 2007. DOI: https://doi.org/10.1007/978-3-540-77255-2_28.
- JUAN, L.; GWUN, O. A comparison of SIFT, PCA-SIFT and SURF. International Journal of Image Processing (IJIP), v. 3, n. 4, p. 143–152, 2009. Disponível em: <https://www.cscjournals.org/manuscript/Journals/IJIP/Volume3/Issue4/IJIP-51.pdf>. Acesso em: 13 mar. 2022.
- LORDELO, J. O significado das tatuagens no mundo do crime e nos presídios. Lordelo Treinamento, 2018. Disponível em: <https://lordellotreinamento.com.br/2018/03/05/o-significado-das-tatuagens-no-mundo-do-crime-e-nos-presidios/>. Acesso em: 10 jan. 2022.
- LOWE, D. G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, v. 60, n. 2, p. 91–110, 2004. DOI: <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- MANGER, D. Large-scale tattoo image retrieval. In: Conference on Computer and Robot Vision. IEEE, 2012. p. 454–459. Disponível em: <https://publicrest.fraunhofer.de/server/api/core/bitstreams/76604e42-93cc-482a-a113-2fea0171218a/content>. Acesso em: 10 jan. 2022.
- MUSCI, M. et al. Drones and Sustainability: How Technology Can Assist in the Automatic Detection of Waste. In Hard-to-Access Areas. Revista De Gestão - RGS, 18(9), e07235. 2024a. DOI: <https://doi.org/10.24857/rgsa.v18n9-109>.
- MUSCI, M. et al. Inteligência artificial e sensoriamento remoto com o uso de vants: uma aplicação na detecção automática de resíduos sólidos urbanos. In: dos Santos, A. F. et al. (Org.). Sustentabilidade e meio ambiente - Inovação e responsabilidade. 1ed. Ponta Grossa: Atena Editora, 2024b, v. 1, p. 81-97. DOI: <https://doi.org/10.22533/at.ed.718242712>.
- MUSCI, M. et al. Uso do algoritmo SIFT para reconhecimento automático de tatuagens. Aracê - Direitos Humanos em Revista, v. 7, p. 10988-11000, 2025. DOI: <https://doi.org/10.56238/arev7n3-053>.
- POCEVIČ, G.; STEFANOVIČ, P.; RAMANAUSKAITĖ, S.; PAVLOV, E. Approach for Tattoo Detection and Identification Based on YOLOv5 and Similarity Distance. Applied Sciences, v. 14, n. 13, p. 5576, 2024. DOI: <https://doi.org/10.3390/app14135576>.
- REDMON, J.; FARHADI, A. YOLO9000: Better, Faster, Stronger. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, Honolulu. Proceedings. Honolulu: IEEE, 2017. p. 6517–6525. DOI: <https://doi.org/10.1109/CVPR.2017.690>.
- SILVA, R. T.; LOPES, H. S. A Transfer Learning Approach for the Tattoo Detection Problem. In: XV Brazilian Congress of Computational Intelligence (SBIC), Online, 3–6 Oct. 2021. DOI: <https://doi.org/10.21528/CBIC2021-34>.
- SUN, Z. H. et al. Tattoo detection and localization using region-based deep learning. In: 23rd International Conference on Pattern Recognition (ICPR), 2016, Cancun. Proceedings. Cancun: IEEE, 2016. p. 3055–3060. DOI: <https://doi.org/10.1109/ICPR.2016.7900103>.

