

Análise de precisão de detectores e descritores em imagens

*Samuel Rebouças de Jesus
Walber Conceição de Jesus Rocha
João Carlos Nunes Bittencourt*

Introdução

Os seres humanos dispõem de recursos intrínsecos que os robôs não apresentam, dentre os quais destaca-se a capacidade de perceber o ambiente a partir de um sensor visual. Ao longo das últimas décadas, técnicas e métodos de simulação de visão computacional em sistemas robóticos têm sido desenvolvidos (ALMEIDA, 2018; CHAO ZHOU; YUCHENG WEI; TIENIU TAN, 2003; COUTO, 2012), permitindo que os mesmos possam extrair informações de modo a entender o ambiente a partir da utilização de câmeras e sensores. No entanto, tais aplicações ainda apresentam limitações, principalmente quando precisam simular funções intrínsecas dos seres humanos. Adicionalmente, com o advento de novas plataformas portáteis e tecnologias de baixo desempenho, torna-se necessária a análise do custo computacional de tais técnicas, tendo em vista subsidiar tomadas de decisão por parte de projetistas de sistemas robóticos.

A visão computacional tornou-se uma tecnologia chave em várias áreas do conhecimento, e tem sido utilizada como elemento central de um vasto número de aplicações, atualmente. Mais especificamente, a extração de características de imagens retrata uma área essencial da visão computacional. Dentro desse contexto, destacam-se os mecanismos de detecção de características (*feature detection*) e descrição de características (*feature description*). Tais abordagens apresentam algoritmos capazes de extrair características

de imagens, transformando dados visuais em informações as quais sistemas computacionais são aptos a manipular (HASSABALLAH; AWAD, 2016).

Um grande conjunto de métodos para detecção, descrição e correspondência de características em imagens tem sido propostos e validados nas últimas décadas (AHMADABADIAN et al., 2013; BAY; TUYTELAARS; VAN GOOL, 2006; CALONDER et al., 2010; LOWE, 1999; RYU; LEE; PARK, 2011). Uma diversidade de aplicações que utilizam desde computadores móveis até sistemas robóticos demandam por mecanismos de detecção de objetos (SCHMIDT; KRAFT; KASIŃSKI, 2010).

Neste sentido, trabalhos que formalizam uma análise sistemática, produzindo dados que expressam a precisão destes algoritmos representa significativa importância, uma vez que aplicações desta natureza não devem desperdiçar recursos obtendo resultados insatisfatórios em decorrência das suas limitações intrínsecas de processamento e complexidade computacional. Uma análise semelhante realizou um estudo acerca da precisão dos algoritmos descritores, determinando o impacto de sua aplicação durante o processo de localização e mapeamento simultâneo do ambiente (VSLAM), não considerando, entretanto, a eficiência geral dos algoritmos estudados (HARTMANN; KLÜSSENDORFF; MAEHLE, 2013).

Este capítulo tem como objetivo apresentar dados que possibilitem realizar uma análise do nível de precisão obtido a partir da combinação de algoritmos detectores e descritores de características quando aplicados em imagens estáticas. Para tanto foi necessário relacionar os dados oriundos da aplicação dos algoritmos, viabilizando assim definir quais as melhores opções de uso em um dado projeto.

O restante deste capítulo é organizado da forma que segue: a Seção 2 descreve os materiais e métodos empregados no desenvolvimento do trabalho; a Seção 3 apresenta os principais

resultados da pesquisa e uma análise dos dados obtidos; as conclusões e perspectivas de trabalhos futuros são apresentadas na Seção 4.

Método

O sistema de software projetado para a realização das análises foi desenvolvido em linguagem de programação Python, versão 2.7, utilizado em conjunto com a biblioteca OpenCV, versão 3.2.0, para implementação das funções referentes à visão computacional e processamento de imagem. Mediante investigação e análise preliminares, foram escolhidos seis algoritmos detectores e cinco algoritmos descritores tais quais apresentados na Tabela 1, resultando assim em um total de trinta combinações entre algoritmos de detecção/descrição de características.

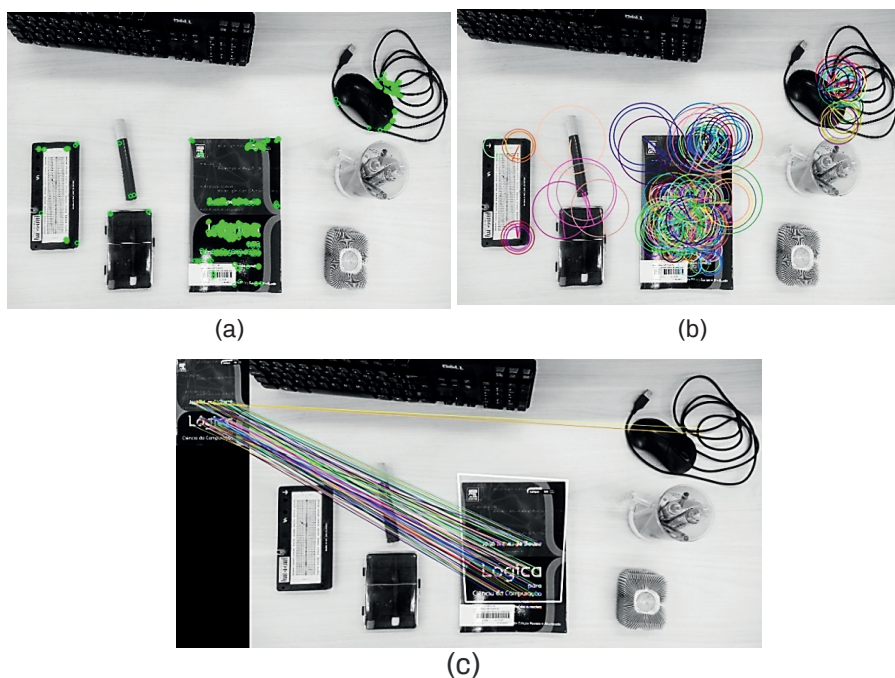
O sistema de extração de característica foi organizado a partir de três componentes principais. O primeiro consiste da utilização do algoritmo detector para obter os pontos de interesse na imagem analisada (*key points*). A Figura 1a expressa uma imagem após a aplicação de um algoritmo detector. Cada círculo, destacado na cor verde, representa um ponto de interesse detectado pelo algoritmo. Posteriormente, aplica-se um algoritmo de descrição de características, representado na Figura 1b. Nesta imagem, cada ponto de interesse é representado por um círculo indicando seu tamanho e orientação.

Tabela 1 –Algoritmos escolhidos para o processo de análise.

| Detector | Descritor |
|------------|-----------|
| FAST | BRIEF |
| Harris | BRISK |
| ORB | FREAK |
| Shi-Tomasi | ORB |
| Star | SIFT |
| SURF | |

Fonte: Autoria Própria (2020).

Figura 1 – Caracterização e correlação de uma imagem. (a) imagem após a aplicação do algoritmo detector. (b) resultado do processo de descrição de características. (c) resultado da correspondência utilizando o método força bruta



Fonte: Autoria Própria (2020).

O presente trabalho visa fornecer informações de precisão associados aos algoritmos de detecção e descrição de características. Deste modo, torna-se necessário realizar a correlação entre uma imagem de referência e um conjunto de imagens de rastreo, com o propósito de obter pontos que apresentem as mesmas características. Para tanto, optou-se pelo uso do método força bruta da biblioteca do OpenCV, denominado BFMatcher (“Feature Matching — OpenCV 3.0.0-dev documentation”, 2014). Este método utiliza os dados referentes ao descritor de uma imagem de referência, relacionando-o com o descritor da imagem de rastreo a ser analisada. A Figura 1c expressa o resultado da utilização do correspondente de força

bruta BFMatcher entre uma imagem de referência (à esquerda) e uma imagem de rastreo. Mediante a utilização dos dados retornados do descritor, é possível obter os pontos correspondentes entre as duas imagens. Com base nesses dados, obtêm-se uma métrica de precisão para um dado algoritmo, a qual pode ser utilizada para medir a sua eficiência geral.

A partir da análise do resultado apresentado na Figura 1c, observa-se que ao redor do objeto de correspondência foi produzido um contorno correspondente à forma geométrica aproximada do mesmo. O método utilizado para este processo é denominado homografia (OPENCV, 2020). Segundo Elan Dubrofsky (2009), a homografia é um mapeamento invertível de pontos e linhas no plano projetivo P^2 . Tal transformação está associada à colineação, projetividade e transformação projetiva planar. Portanto, ao passar o conjunto de pontos definidos como correspondências entre as imagens, este processo definirá a transformação de perspectiva, sendo possível definir, ao menos, quatro pontos de referência deste objeto. Tal processo permite que seja traçado o contorno do mesmo na forma de uma representação poligonal. O processo de homografia, portanto, possibilita utilizar os dados obtidos pela dupla detector/descritor para a localização e rastreo de objetos em uma imagem, ou em um fluxo de imagens.

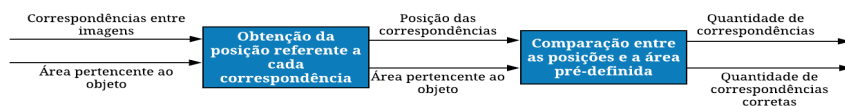
Análise de precisão dos algoritmos

A análise de precisão dos algoritmos tem por finalidade aferir a qualidade dos resultados obtidos a partir da combinação dos algoritmos descritores e detectores. O conjunto de imagens adotado neste trabalho foi analisado tendo em vista obter as relações de pontos semelhantes entre elas. Dessa forma, a quantidade de pontos semelhantes obtidos, juntamente com a informação acerca

de quantos destes pontos foram detectados corretamente, conduzem a uma análise quantitativa dos resultados de precisão associados a cada combinação de algoritmos.

A partir das correspondências obtidas por meio do sistema de extração de características, a definição da precisão dos algoritmos acrescenta apenas a análise referente a posição das correspondências. O fluxo descrito na Figura 2 consiste na verificação de cada correspondência obtida, e na determinação referente à área onde o objeto está localizado.

Figura 2—Diagrama de representação do teste de precisão no sistema de análise.



Fonte: Autoria Própria (2020).

Para cada imagem utilizada durante os testes, a região a qual o objeto de referência se encontra é diferente. Devido a essa característica, tais regiões foram previamente identificadas e fornecidas como parâmetro de entrada para o sistema de análise. Quando os pontos relacionados expressam a mesma região, tal correspondência é computada como correta (match). Ao final deste processo, observa-se a quantidade de correspondências obtidas e o total de correspondências corretas obtidas em cada uma das treze imagens pertencentes ao conjunto. Tendo em vista realizar uma análise justa dos algoritmos, esse conjunto de imagens não levou em consideração aspectos de variância, tais como rotação, distância e luminosidade.

Em detrimento das distinções operacionais de cada algoritmo, a quantidade de correspondências obtidas durante os testes apresentava uma média de distribuição desigual, dada em razão

dos diferentes métodos de obtenção dos pontos chave. Deste modo, julgou-se necessário limitar a quantidade de *matches* analisados para cada par de algoritmos até um máximo de 150 correspondências. Deste modo não houve interferência na análise dos algoritmos que apresentavam uma quantidade elevada de correspondências.

Resultados e discussões

Os testes realizados ao longo desta pesquisa relacionaram dados referentes à quantidade de correspondências obtidas e correspondências corretas, apresentados na Tabela 2. Além disso, o estudo permitiu extrair a capacidade de rastreamento de objetos, dada a partir do total de correspondências corretas entre as imagens testadas. Entretanto, analisar apenas os dados de identificação e rastreamento isoladamente não é suficiente para realizar uma análise contundente a respeito da eficiência geral das combinações de algoritmos. Diante disso, relacionaram-se os dados de forma a produzir valores percentuais referentes à relação entre a quantidade de correspondências obtidas e correspondências corretas. Tais informações são expressas na Tabela 3.

Tabela 2–Quantidade de correspondências obtidas (T) e correspondências corretas (A).

| | BRIEF | | BRISK | | FREAK | | ORB | | SIFT | |
|-------------------|-------|-----|-------|-----|-------|-----|-----|-----|------|-----|
| | T | A | T | A | T | A | T | A | T | A |
| FAST | 150 | 139 | 150 | 95 | 150 | 115 | 150 | 123 | 150 | 141 |
| Harris | 19 | 15 | 29 | 14 | 24 | 15 | 19 | 14 | 35 | 14 |
| ORB | 150 | 141 | 145 | 135 | 93 | 92 | 150 | 138 | 150 | 139 |
| Shi-Tomasi | 39 | 29 | 49 | 34 | 40 | 27 | 36 | 25 | 48 | 33 |
| Star | 45 | 32 | 47 | 25 | 42 | 28 | 44 | 26 | 46 | 37 |
| SURF | 150 | 110 | 150 | 117 | 150 | 107 | 150 | 96 | 150 | 142 |

Fonte: Autoria Própria (2020).

O padrão de distribuição da quantidade de acertos obtidos a partir da junção dos algoritmos expressa influência direta dos algoritmos detectores. Quanto mais pontos de interesse são detectados, maiores são as chances de obter a correlação entre uma imagem de referência e imagem de rastreo. Entretanto, a influência dos algoritmos descritores nesse aspecto também deve ser considerada. A descrição de cada ponto de interesse produz informações sobre o mesmo. Além disso, quanto mais evidentes forem essas informações, maior a chance de pontos semelhantes serem relacionados.

A influência dos algoritmos descritores pode ser identificada analisando a Tabela 2, mais especificamente, a linha referente ao algoritmo detector ORB. Ao ser utilizado em conjunto com os algoritmos descritores BRIEF, BRISK, ORB e SIFT, a quantidade de acertos obtidos foi igual ou próxima do limite estipulado. Ao utilizar este algoritmo detector em conjunto com o descritor FREAK, a quantidade de acertos foi relativamente menor, denotando uma seletividade na correlação dos pontos de interesse em comparação com as outras combinações. Tal informação é evidenciada a partir da análise da Tabela 3, que expõe 99% de acerto das correspondências obtidas para esta combinação.

Na Tabela 3, a maior parte dos algoritmos avaliados apresentaram uma porcentagem de acerto para as correspondências detectadas elevada. Entretanto, este valor não representa apenas um dado quantitativo, uma vez que quanto maior a porcentagem de correspondências corretas detectadas entre duas imagens, maiores também serão as chances do objeto rastreado ser detectado, denotando assim a qualidade dos pontos obtidos na detecção.

Tabela 3–Percentual de correspondências corretas dos pares de algoritmo.

| | BRIEF | BRISK | FREAK | ORB | SIFT |
|-------------------|--------------|--------------|--------------|------------|-------------|
| FAST | 89,54% | 63,54% | 76,72% | 82,21% | 94,05% |
| Harris | 76,83% | 48,18% | 63,01% | 73,64% | 41,18% |
| ORB | 94,62% | 93,04% | 99,00% | 92,15% | 93,18% |
| Shi-Tomasi | 75,05% | 69,20% | 69,35% | 69,15% | 68,89% |
| Star | 71,09% | 54,34% | 67,03% | 60,35% | 81,00% |
| SURF | 73,59% | 78,05% | 71,54% | 64,00% | 94,92% |

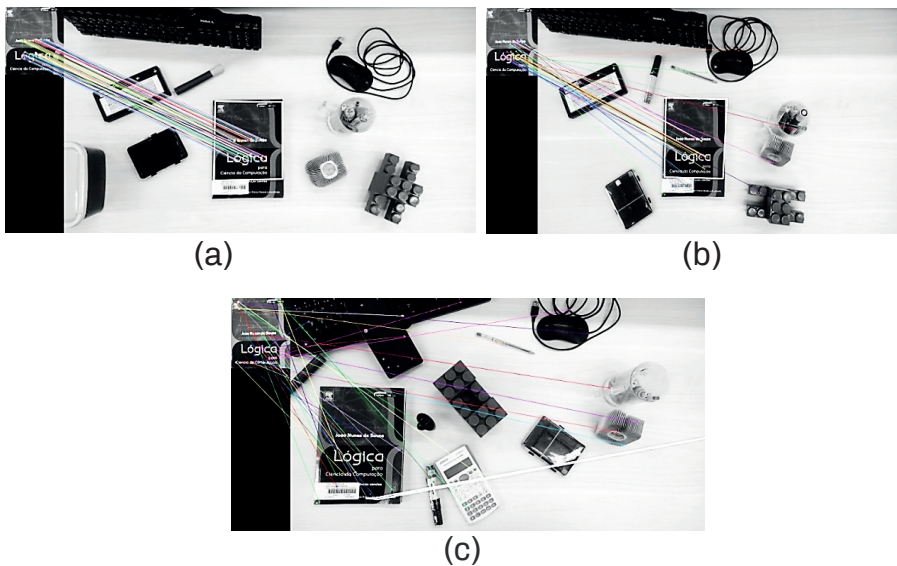
Fonte: Autoria Própria (2020).

Ao avaliar os resultados apresentados nas imagens após o processo de correlação das mesmas, foi verificado que 60% dos algoritmos apresentaram aproveitamento acima de 70% ao detectar o objeto nas diferentes imagens utilizadas. Por outro lado, os algoritmos oriundos da junção do detector Harris com descritor BRISK e detector Harris com descritor SIFT, por exemplo, apresentaram um percentual abaixo da média, apresentando assim os resultados menos eficientes em relação às taxas de acerto para as correspondências detectadas.

Durante o processo de análise, foram utilizadas treze imagens para cada par de algoritmos, de forma tal que cada imagem evidencie as correspondências obtidas. A Figura 3a representa o resultado de uma imagem analisada pelo algoritmo detector ORB em conjunto com o algoritmo descritor FREAK. Verifica-se que as correspondências obtidas estão concentradas no objeto alvo da análise, corroborando o percentual de precisão referido na Tabela 3. Mesmo esta mesma combinação de algoritmos não alcançando o número máximo de correspondências estabelecida como métrica de análise, as mesmas denotaram satisfatória qualidade, demonstrando que a quantidade de correspondências obtidas e a assertividade das mesmas não necessariamente são grandezas diretamente proporcionais.

A Figura 3b corresponde à combinação dos algoritmos Star e BRIEF. Pode-se notar, neste caso específico, a forma com que as correspondências obtidas estão dispersas, uma vez que este par apresentou 71,09% de acerto em relação às correspondências detectadas. No entanto, observa-se que, apesar da captura das falsas correspondências, o objeto foi selecionado corretamente, não afetando assim a detecção do mesmo. Uma vez que a homografia foi bem-sucedida, a utilização de algoritmos com essa taxa de precisão possibilita o desenvolvimento de aplicações de rastreo.

Figura 3—Amostra de resultados obtidos durante o teste de precisão.



Fonte: Autoria Própria (2020).

Por fim, a Figura 3c representa a combinação entre os algoritmos Harris e SIFT. Observa-se que as correspondências estão, em sua maioria, dispersas e equivocadas, corroborando com o percentual de 41,18% de correspondências corretas, indicado na Tabela 3. Devido à baixa qualidade dos dados obtidos, a identificação correta do objeto foi comprometida.

Conclusão

Nas aplicações de visão computacional, a precisão é um fator determinante no sentido de expressar a eficiência de um dado sistema. À luz de tal afirmação, ser capaz de analisar, extrair características de imagens e correlacionar padrões representa demasiada importância, uma vez que a definição das coordenadas de um dado objeto no ambiente pode ser determinada como uma informação crítica.

O presente trabalho apresentou uma análise da eficiência dos principais algoritmos descritores e detectores de características presentes na literatura, e como a combinação de entre eles exerce influência na eficiência de um sistema. A partir dos dados obtidos ao longo da referida análise, constatou-se a influência que os algoritmos detectores exercem sobre a quantidade de correspondências capturadas. Mediante avaliação de eficiência, foram definidas combinações de algoritmos que apresentam os melhores resultados em relação à precisão. Neste sentido, as combinações obtidas a partir do detector ORB e as combinações entre detector/descritor SURF/SIFT e FAST/SIFT apresentaram resultados que expressam uma qualidade superior em relação às demais combinações.

A principal limitação imposta durante o processo de análise caracteriza-se pela ausência de variância em relação a rotação, luminosidade e distância nas imagens utilizadas. Tal restrição foi definida tendo em vista uma análise justa dos algoritmos, uma vez que nem todos caracterizam-se por serem sensíveis a tais variâncias. Deste modo, a análise de precisão e desempenho dos algoritmos é restrita a um cenário “ideal”, no qual a câmera e o objeto estejam sempre perpendiculares, à uma distância favorável e em boas condições de luminosidade.

A partir do desenvolvimento do trabalho apresentado neste capítulo, vislumbra-se a construção de plataformas de rastreo de objetos nas quais propriedades como rotação, luminosidade,

distância e ruído não influenciem no processo de correlação das imagens. Além disso, torna-se evidente uma avaliação criteriosa dos algoritmos selecionados, considerando-se correspondências sensíveis à rotação e outras não idealidades.

Referências

AHMADABADIAN, A. H. et al. ***A comparison of dense matching algorithms for scaled surface reconstruction using stereo camera rigs***. ISPRS Journal of Photogrammetry and Remote Sensing, v. 78, p. 157–167, 1 abr. 2013.

ALMEIDA, A. F. DE. **Implementação de algoritmo de visão computacional para comunicação entre membros de um enxame de robôs**. 2018. 43 f. Trabalho de Conclusão de Curso (Graduação em Engenharia Mecatrônica) - Universidade Federal de Uberlândia, Uberlândia, 2018.

BAY, H.; TUYTELAARS, T.; VAN GOOL, L. SURF: ***Speeded Up Robust Features***. (A. Leonardis, H. Bischof, A. Pinz, Eds.) Computer Vision – ECCV 2006. Anais...: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006

CALONDER, M. et al. BRIEF: ***Binary Robust Independent Elementary Features***. (K. Daniilidis, P. Maragos, N. Paragios, Eds.) Computer Vision – ECCV 2010. Anais...: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2010

CHAO ZHOU; YUCHENG WEI; TIENIU TAN. ***Mobile robot self-localization based on global visual appearance features***. 2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422). Anais... In: 2003 IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION (CAT. NO.03CH37422). set. 2003

COUTO, L. N. **Sistema para localização robótica de veículos autônomos baseado em visão computacional por pontos de**

referência. Mestrado em Ciências de Computação e Matemática Computacional—São Carlos: Universidade de São Paulo, 18 maio 2012.

ELAN DUBROFSKY. *Homography Estimation. Homography Estimation*, 2009.

_____. *FeatureMatching—OpenCV3.0.0-devdocumentation*. Disponível em: <https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_matcher/py_matcher.html>. Acessoem: 15 jun. 2020.

HARTMANN, J.; KLÜSSENDORFF, J. H.; MAEHLE, E. **A comparison of feature descriptors for visual SLAM**. 2013 European Conference on Mobile Robots. Anais... In: 2013 EUROPEAN CONFERENCE ON MOBILE ROBOTS. set. 2013

HASSABALLAH, M.; AWAD, A. I. *Detection and Description of Image Features: An Introduction*. In: AWAD, A. I.; HASSABALLAH, M. (Eds.). . Image Feature Detectors and Descriptors: Foundations and Applications. Studies in Computational Intelligence. Cham: Springer International Publishing, 2016. p. 1–8.

LOWE, D. G. *Object recognition from local scale-invariant features*. Proceedings of the Seventh IEEE International Conference on Computer Vision. Anais... In: PROCEEDINGS OF THE SEVENTH IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION. set. 1999

OPENCV. *OpenCV: Basic concepts of the homography explained with code*. Disponível em: <https://docs.opencv.org/master/d9/dab/tutorial_homography.html>. Acessoem: 15 jun. 2020.

RYU, J.-B.; LEE, C.-G.; PARK, H.-H. *Formula for Harris corner detector*. Electronics Letters, v. 47, n. 3, p. 180–181, fev. 2011.

SCHMIDT, A.; KRAFT, M.; KASIŃSKI, A. *An Evaluation of Image Feature Detectors and Descriptors for Robot Navigation*. In: BOLC,

L. et al. (Eds.). . Computer Vision and Graphics. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. v. 6375p. 251–259.