

Novo Algoritmo para Reconhecimento de Objetos Invariante Afim

Guillermo Ángel Pérez López e Hae Yong Kim

Resumo — Recentemente, um dos autores propôs um algoritmo de reconhecimento de objetos (casamento de modelos) invariante à rotação e escala chamado Forapro. Este algoritmo utiliza coeficientes de Fourier de projeções radiais e circulares para obter características locais invariantes à rotação. Neste trabalho, tornamos Forapro invariante à distorção de perspectiva utilizando a técnica de simulação de pontos de vista. Este novo algoritmo, que denominamos de Aforapro, é robusto a padrões repetidos e mudanças de brilho e contraste. Aforapro foi testada com 350 imagens e demonstrou ser bastante robusto a diferentes distorções.

Palavras-Chave — Reconhecimento de objetos, casamento de modelos, invariância a perspectiva, coeficientes de Fourier, transformação afim, ASIFT, Forapro.

Abstract — Recently one of the authors proposed an object recognition (template matching) algorithm invariant to rotation and scaling named Forapro. This algorithm uses the Fourier coefficients of radial and circular projections to compute rotation-invariant local features. In this paper, we make Forapro invariant to perspective distortion using the viewpoint simulation technique. This new algorithm, that we named Aforapro, is robust to repetitive patterns and changes in brightness and contrast. Aforapro was tested with 350 images and has proved to be quite robust to different distortions.

Keywords — Object recognition, template matching, perspective invariance, Fourier coefficients, affine transformation, ASIFT, Forapro.

I. INTRODUÇÃO

O casamento de modelos (também chamado de reconhecimento de objetos) consiste em achar as instâncias de uma imagem modelo Q (*query*) presentes numa outra imagem A (*analyzed*). Neste artigo, consideraremos somente as imagens em níveis de cinza. A figura 1 ilustra um exemplo de casamento de modelos. Um padrão é um pedaço de Q que pode ser utilizado para localizar Q em A .

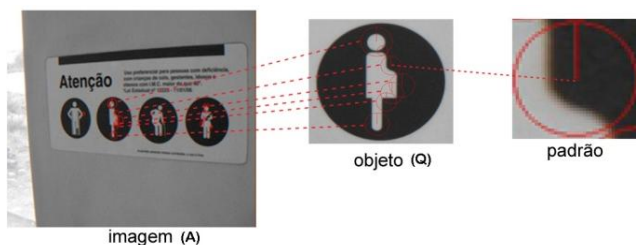


Fig. 1. Processo do reconhecimento de objetos

Algumas das dificuldades presentes neste processo são: (1) as imagens podem aparecer em diferentes rotações e escalas;

(2) as instâncias de Q presentes em A podem ter sofrido efeitos fotométricos que incluem mudança de brilho e contraste; (3) o conjunto ótico da câmera pode borrar as imagens (imagem fora de foco); (4) alguns padrões de Q (subpartes de Q) podem aparecer repetidamente em A fora das instâncias de Q ; (5) as imagens podem sofrer deformação em perspectiva devido à mudança na posição da câmera.

O nosso objetivo é propor um algoritmo de reconhecimento de objetos invariante a deformações em perspectivas, causadas pela variação do ponto de vista da câmera.

Uma solução para obter esta invariância é utilizar descritores locais invariantes à transformação afim [1], pois (como veremos adiante) uma deformação em perspectiva pode ser aproximada localmente por uma transformação afim. Esta abordagem não apresenta resultados satisfatórios, em parte devido a não comutabilidade entre a deformação em perspectiva e o borrão introduzido pelo conjunto ótico.

Outra solução é utilizar a “simulação de pontos de vista”. Esta técnica foi proposta por Morel e Yu [2] para estender o popular SIFT (*scale invariant feature transform*) [3] a fim de torná-lo invariante afim, obtendo o algoritmo ASIFT (*affine-SIFT*). A robustez obtida pelo ASIFT o destaca como um dos melhores algoritmos invariantes afins no estado da arte. Porém o seu desempenho é fraco na presença de padrões repetidos ou grandes mudanças de brilho e contraste.

Recentemente, um dos autores deste artigo propôs um algoritmo de casamento de modelos invariante à rotação e escala chamado Forapro (*Fourier coefficients of Radial Projections*) [4]. Forapro utiliza os coeficientes de Fourier de projeções radiais e circulares para obter características locais invariantes à rotação. Neste trabalho, tornamos Forapro invariante às transformações afins utilizando a técnica de simulação de pontos de vista, obtendo o algoritmo que denominamos de Aforapro (Affine-invariant Forapro). Testamos Aforapro usando imagens com diferentes deformações e o comparamos com o ASIFT.

Este artigo está organizado como segue: a Seção 2 apresenta os conceitos das características invariantes; a Seção 3 apresenta o novo algoritmo chamado Aforapro; a Seção 4 apresenta os resultados experimentais; e a Seção 5 apresenta as conclusões.

II. CARACTERÍSTICAS INVARIANTES

A. Transformação afim

A transformação afim de uma imagem I_1 é um mapeamento que a leva a outra imagem I_2 :

$$I_2(x_2, y_2) = A * I_1(x_1, y_1) + T \tag{1}$$

onde A é uma matriz 2x2 inversível e $T \in R^2$ é o vetor de translação. A transformação afim pode ser decomposta:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & -\text{sen} \phi \\ \text{sen} \phi & \cos \phi \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \tag{2}$$

onde S_x e S_y são os fatores de escala para os eixos x e y respectivamente, k é um fator de não ortogonalidade entre eixos, ϕ é o ângulo de rotação sobre o plano, e e f são fatores de translação. Resumindo, a transformação afim de uma imagem é uma transformação linear com seis graus de liberdade (2 mudanças de escala, 2 rotações e 2 translações):

$$I(x, y) \rightarrow I(ax + by + e, cx + dy + f) \tag{3}$$

As propriedades de uma transformada afim no espaço euclidiano são:

- 1) Preserva a relação de colinearidade entre pontos, isto é, três pontos colineares continuam sendo colineares após a transformação. Como consequência desta propriedade, a transformação afim preserva o paralelismo das retas.
- 2) Preserva a razão das distâncias ao longo de uma linha. Isto é, para três pontos colineares distintos p_1, p_2 e p_3 , a razão $|p_2 - p_1|/|p_3 - p_2|$ é preservada.

B. Invariância a pontos de vista

O ponto de vista é a posição espacial do fotógrafo em relação à cena, no instante da captura de imagem. Um objeto pode ser fotografado de múltiplos pontos de vista. A mudança do ponto de vista da câmera introduz uma deformação em perspectiva nos objetos que pode dificultar o reconhecimento do objeto. Alguns algoritmos procuram a invariância a pontos de vista utilizando a seguinte propriedade:

Propriedade: a aparente deformação em perspectiva de um objeto, induzida pela mudança de ponto de vista, pode ser aproximada localmente por uma transformação afim [2, 1].

Em nosso trabalho adotamos a técnica de simulação de pontos de vista, esta técnica baseia-se na utilização dos parâmetros de orientação da câmera do seguinte modelo afim:

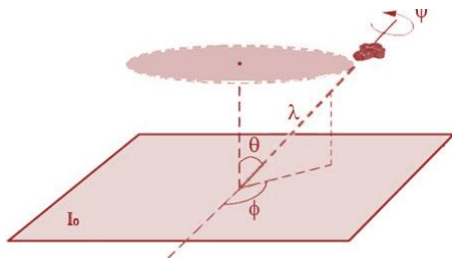


Fig. 2. Parâmetros da captura da imagem. Figura modificada de [2].

Na Fig. 2, o plano I_0 é a vista frontal do objeto físico, ψ é o ângulo de rotação da câmera sobre seu eixo ótico, λ é o parâmetro de escala (a distância câmera ao ponto central do plano), ϕ é o ângulo de rotação da câmera sobre o plano da

imagem, e θ é o ângulo de inclinação medido entre a normal do plano da imagem e o eixo ótico da câmera. Assume-se que a vista frontal é determinada por $\lambda = 1$ e $\theta = \phi = \psi = 0$.

Este modelo transforma o núcleo da Equação 1 em:

$$A = H_\lambda R_1(\psi) T_t R_2(\phi) = \lambda \begin{bmatrix} \cos \psi & -\text{sen} \psi \\ \text{sen} \psi & \cos \psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & -\text{sen} \phi \\ \text{sen} \phi & \cos \phi \end{bmatrix} \tag{4}$$

onde H_λ é a matriz de escala com $\lambda > 0$, os R_i são as rotações, e T_t é a mudança de inclinação com $t = 1/\cos \theta$. O parâmetro t é chamado de *tilt* (grau de inclinação de uma vista para outra) e pode ser: tilt absoluto que é a diferença de inclinação de uma imagem para sua vista frontal, e tilt de transição que é a medida de inclinação entre dois pontos de vistas diferentes de uma imagem. Em nosso trabalho utilizamos somente o tilt absoluto. Isto é, assumiremos que o modelo Q foi fotografado de um ponto de vista aproximadamente frontal.

A simulação de pontos de vista varia os valores de θ e ϕ para simular todas as possíveis distorções causadas pela mudança de orientação da câmera. A escolha correta do intervalo de valores destes dois parâmetros garante que um pequeno número de imagens simuladas seja o suficiente para ter correspondência com todas as distorções presentes nas imagens capturadas. A Fig. 3 ilustra algumas dessas simulações. Note que dos seis graus de liberdade da transformação afim, estamos simulando apenas dois. Foraproprio irá cuidar dos quatro restantes, por ser invariante à rotação, escala e translação.

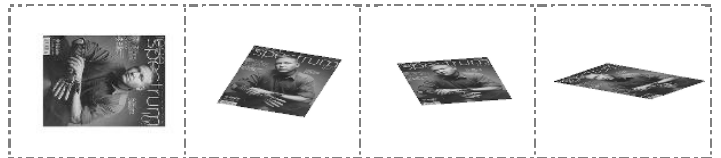


Fig. 3. Algumas das simulações de pontos de vista.

C. Invariância a mudanças de brilho e contraste

Definimos que duas imagens I_1 e I_2 são equivalentes sob mudança de brilho e contraste se existir um fator de correção de contraste $\beta > 0$ e outro fator de correção de brilho γ tais que $I_2 = \beta I_1 + \gamma I$, onde I é a matriz de 1's. Sejam imagem modelo Q e imagem a analisar A . Consideramos que um algoritmo de reconhecimento de objetos é invariante a brilho e contraste se consegue achar instâncias de Q em A independentemente do brilho e contraste da instância.

D. Robustez a padrões repetidos

Uma imagem de cenas reais pode conter muitos elementos que se repetem (e.g. caracteres na escrita ou azulejos no chão). Neste caso, o formato do objeto procurado pode ser similar à estrutura do fundo ou similar a outros objetos presentes na cena. Estes padrões repetidos criam dificuldades no processo de casamento de padrões, aumentando o número de casamentos errados, dado que eles são localmente indistinguíveis [5].

Sejam uma imagem modelo Q e uma imagem a analisar A , onde n instâncias de Q aparecem dentro de A . Dizemos que A

possui padrões de Q repetidos se algumas partes de Q aparecerem em A fora das n instâncias de Q . Definiremos que um algoritmo de reconhecimento de objetos é robusto a padrões repetidos se consegue achar as n instâncias de Q em A , mesmo que A possua padrões de Q repetidos.

III. ALGORITMO AFORAPRO

Dadas duas imagens Q e A , assumiremos que a imagem Q apresenta o ponto de vista frontal e a imagem A contém uma instância de Q fotografada a partir de um ponto de vista aleatório. Diferentemente do algoritmo ASIFT, não consideramos as “inclinações de transição” onde ambas Q e A podem estar inclinadas em direções diferentes. Esta suposição não é muito restritiva na prática, pois normalmente o usuário possui controle sobre o processo de aquisição do modelo Q . Então, os passos do algoritmo Aforapro podem ser resumidos:

1. Transforma-se a imagem Q , simulando as possíveis distorções causadas pela mudança de ponto de vista da câmera, para obter as imagens sintéticas ($Q \rightarrow Q_1, Q_2, \dots, Q_n$). Supomos que a imagem Q foi fotografada na posição frontal, mas que pode aparecer na imagem A com distorção. Diferentemente do algoritmo ASIFT, não consideramos as “inclinações de transição” onde ambas Q e A podem estar inclinadas em direções diferentes. As distorções dependem dos parâmetros de rotação ϕ e inclinação $\theta \rightarrow t = 1/\cos \theta$.
2. Cada imagem simulada Q_i é procurada na imagem A , com o objetivo de escolher o modelo Q_i que apresente o maior grau de semelhança à instância da imagem Q presente na imagem A .
3. Realiza-se o casamento de padrões entre o modelo Q_i escolhido e a imagem A , localizando e rotulando espacialmente a instância encontrada.

Para dar início à execução do Aforapro deve-se definir o número de simulações, este parâmetro assegura que a quantidade das imagens simuladas seja suficiente para cobrir todos os possíveis pontos de vista que a câmera pode adotar. O número de simulações tem dependência direta com a variação dos parâmetros de rotação e inclinação dos eixos da câmera. Seguindo a sugestão de Morel e Yu [7] os valores adotados são $\Delta t = t_{k+1}/t_k = \sqrt{2}$ e $\Delta \phi = 72^\circ/t$. Em consequência, os intervalos são:

- $t = \{\sqrt{2}^0, \sqrt{2}^1, \sqrt{2}^2, \sqrt{2}^3, \sqrt{2}^4\}$, onde $t = 1/\cos \theta$
- $\phi = \{0, b/t, \dots, kb/t\}$, onde $b = 72^\circ$ e $kb/t < 180^\circ$.

Assim, Aforapro gera 27 imagens simuladas Q_i e atinge uma inclinação de até $\theta \approx 76^\circ$. Mas a quantidade de imagens simuladas pode ser alterada de acordo com a aplicação, por exemplo, $t = \sqrt{2}^5$ que gera 41 Q_i e atinge um intervalo de inclinação até $\theta \approx 80^\circ$. O Aforapro obtém a invariância a pontos de vista pela técnica de “simulação de pontos de vista”

adotada do ASIFT e a robustez a padrões repetidos e mudanças de contraste é herdada do Forapro [6, 4].

IV. RESULTADOS

Para avaliar o algoritmo proposto, construímos uma base de dados com 350 imagens divididas em cinco conjuntos. Cada conjunto foi subdividido em grupos que continham basicamente um objeto modelo Q e várias imagens onde aparece uma instância de Q fotografada desde pontos de vista diferentes (Fig. 4).

Conjunto 1: Composto por 100 imagens divididas em 10 grupos. Fotografamos objetos de uso comum (martelo, alicate, controle de televisão, revista, flores, chinelo, pente, colher, óculos e extensão elétrica) numa sala de laboratório.

Conjunto 2: Composto por 100 imagens de objetos (placas, anúncios, construções, logotipos, grafites e aparelhos) divididas em 20 grupos. Os cenários foram ambientes externos.

Conjunto 3: Composto por 50 imagens de uso frequente na área do casamento de modelos invariante afim, obtidas dos bancos de dados de Morel e Yu [7] e Mikolajczyk [8]. Estas imagens têm sido frequentemente usadas como ponto de referência por outros pesquisadores, inclusive para avaliar o ASIFT.

Conjunto 4: Composto por 50 imagens de objetos de uso comum divididas em 10 grupos. Uma alteração de contraste foi induzida nestas imagens realizando mudanças de fundo, variação de iluminação e efeitos de sombras. O cenário foi uma sala de laboratório.

Conjunto 5: Composto por 50 imagens divididas em 10 grupos. Neste conjunto utilizaram-se imagens de cenas de ambientes naturais que apresentam padrões repetidos (sinais, anúncios, silhuetas, arbustos e aparelhos).

O objetivo é procurar o modelo Q nas imagens onde ele aparece. Foram usados dois critérios de avaliação, pois os algoritmos destinados ao casamento de padrões (e.g. ASIFT) usam linhas de correspondência para indicar os casamentos encontrados e os algoritmos destinados exclusivamente ao reconhecimento de objetos (e.g. Aforapro) utilizam um rótulo sobre o objeto reconhecido. Considera-se o casamento bem sucedido nas seguintes condições: (a) Aforapro, se o traço do retângulo fica no interior do objeto ou se ao menos 50% da área do mesmo fica dentro do objeto procurado; (b) ASIFT, se ao menos 30% das linhas de correspondência marcam casamentos corretos. O tamanho médio das imagens de nossa base de dados é 650x488 pixels.

Foram realizadas três classes de experimentos, utilizando o Aforapro e ASIFT. Os resultados demonstram que o Aforapro apresenta invariância a pontos de vista, e robustez a padrões repetidos e a mudança de brilho e contraste.

A. Experimentos de invariância a pontos de vista

Na realização destes testes utilizaram-se as imagens dos conjuntos 1, 2 e 3. A versão do algoritmo ASIFT utilizada foi a disponibilizada pelos autores [5] e a plataforma de execução foi Windows Vista. O Gráfico 1 apresenta os resultados, pode-

se perceber que Aforapro supera ao ASIFT nesta aplicação. A resposta do ASIFT é boa para objetos que apresentam abundantes detalhes na estrutura (revista, flores), mas tem dificuldades para reconhecer objetos de formas simples e pouca textura (óculos, martelo).

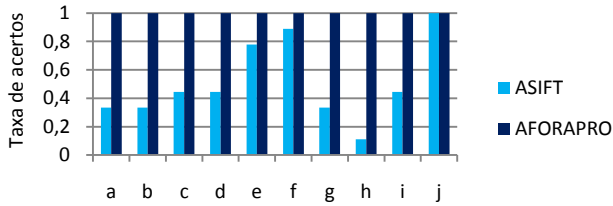


Gráfico 1. Taxas de acertos usando o conjunto 1. a –alicate, b –chinelos, c – controle de tv, d –colher, e –extensão elétrica, f –flores, g –martelo, h –óculos, i –pente, j –revista.

O Gráfico 2 apresenta as taxa de acertos para cada grupo de objetos do Conjunto 2. Verifica-se que a resposta do ASIFT é boa nos grupos de objetos que apresentam abundantes detalhes na estrutura (revista, grafite) e medíocre para grupos onde os objetos possuem poucos detalhes (logotipos).

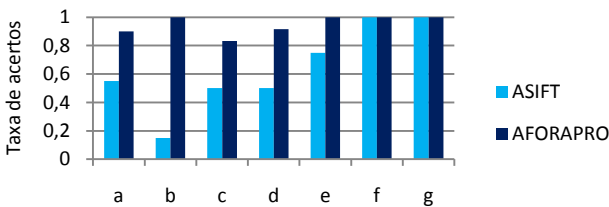


Gráfico 2. Taxas de acertos usando o conjunto 2. a –anúncios, b –objetos tipo A, c –logotipos, d –sinais de trânsito, e – objetos tipo B, f –revista, g –grafite.

O Gráfico 3 apresenta as taxas de acertos para os objetos do Conjunto 3. Estes resultados permitem-nos verificar a invariância a pontos de vista do Aforapro. Os testes realizados neste experimento, além de permitir comparar o Aforapro com o ASIFT, também nos permitem uma comparação indireta do Aforapro com os três outros métodos invariantes a transformações afins (M-SER, Harris-affine, Hessian-affine [1]), pois nos resultados obtidos no trabalho de Morel e You utilizando as mesmas imagens conclui-se que nenhum destes métodos consegue achar objetos com mudança de ponto de vista $\theta \geq 70^\circ$.

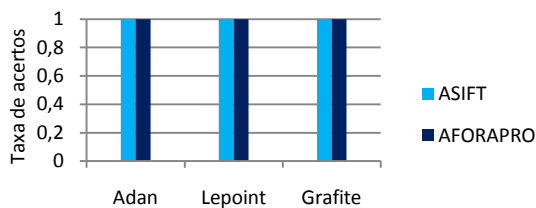


Gráfico 3. Taxas de acertos usando o Conjunto 3.

B. Experimentos de mudanças de brilho e contraste

Na realização destes experimentos foram utilizadas as imagens do Conjunto 4. O Gráfico 4 apresenta as taxas de acertos obtida pelos algoritmos. O desempenho do Aforapro supera amplamente o ASIFT. O ASIFT tem um bom

desempenho no reconhecimento de objetos que apresentam abundância de detalhes (e.g. revista).

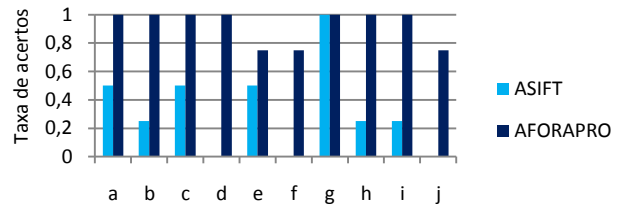


Gráfico 4. Taxas de acertos usando o conjunto 4. a –angulômetro, b – calculadora, c- caneca, d –celular, e –separador, f –conector T, g –revista, h – chave, i –USB, j –brinquedo.

Estes resultados mostram que o Aforapro é invariante a pontos de vista e à mudança de contraste, e que seu desempenho supera o ASIFT.

C. Experimentos de robustez a padrões repetidos

Na realização destes experimentos foram utilizadas as imagens do conjunto 5. O Gráfico 5 apresenta as taxas de acertos. Há dois casos onde Aforapro acerta 100% (informação e elefante) e ASIFT acerta 0%. Nestes experimentos, ASIFT manteve a tendência, obtendo bom desempenho nos objetos que possuem mais detalhes na estrutura.

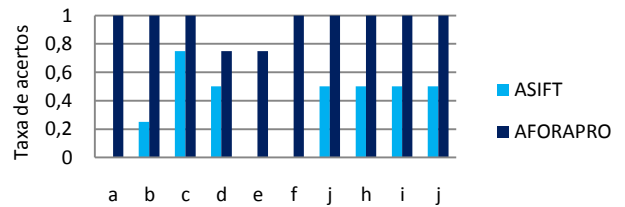


Gráfico 5. Taxas de acertos usando conjunto 5. a –informação, b –arbusto, c –cálcio, d –carro, e –prendedor, f –elefante, g HD, h –face, i –letra, j –deficiente.

D. Resumo dos Resultados

ASIFT foi utilizado como referência para demonstrar a qualidade do desempenho de Aforapro nas três classes de experimentos realizados. O Gráfico 6 apresenta uma síntese das taxas de acertos obtidas em cada experimento.

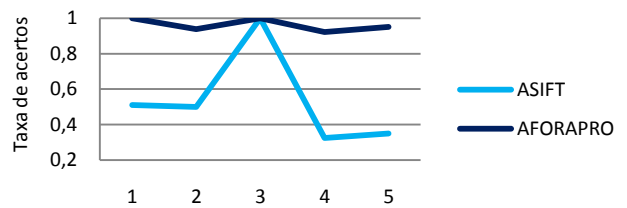


Gráfico 6. Resumo dos resultados obtidos.

O Gráfico 6 mostra que em todos os casos Aforapro supera a marca dos 90% na taxa de acerto. No entanto, ASIFT tem seu pior desempenho nos conjuntos 4 e 5, que são aqueles que possuem variação de brilho e contraste e padrões repetidos.

V. CONCLUSÕES

Neste artigo, propusemos um algoritmo de reconhecimento de objetos invariante a pontos de vista e robusto à mudança de

brilho e contraste, e padrões repetidos que denominamos de Aforapro. Foram realizadas três classes de experimentos (invariância a pontos de vista, robustez a mudanças de contraste e a padrões repetitivos) que nos permitiram avaliar o desempenho do Aforapro. Os resultados demonstraram a superioridade do Aforapro (taxa de acerto 96,20%) contra ASIFT (taxa de acerto 53,72%). Estes resultados são relevantes levando a conta que o ASIFT é atualmente o melhor método da literatura para o casamento de padrões invariante a pontos de vista.

ASIFT demonstrou em todos os experimentos ser aproximadamente quatro vezes mais rápido do que Aforapro. Os tempos no desempenho do Aforapro em cada um dos testes tiveram baixa variação. Em média o Aforapro gastou 8,52 minutos em cada teste. No entanto, a variação dos tempos no ASIFT foi mais significativa, resolvendo rapidamente aqueles casos onde não apresentou casamentos (0,56 minutos em média), mantendo uma média de 2,56 minutos nos casos onde conseguiu algum tipo de reconhecimento e gastando muito mais tempo (9,50 minutos em média) nas imagens que apresentavam quantidade elevada de detalhes na estrutura (e.g. revistas, grafites). A média geral no desempenho do ASIFT foi 2,35 minutos. A velocidade de processamento do Aforapro poderia ser melhorada substancialmente colocando uma estrutura de dados apropriada para indexar os vetores de características. Aforapro é constituído por 3 etapas (gerar Q_i , escolha do Q_i apropriado e execução do casamento) e 70% do tempo de execução é consumido na segunda etapa. Para o

futuro, queremos melhorar o desempenho computacional do Aforapro utilizando uma estrutura de dados para acelerar o processo de recuperação das correspondências entre os vetores de características.

REFERÊNCIAS

- [1] K. Mikolajczyk, et. al, A comparison of affine region detector. International journal of computer vision. vol. 65 (1). pp. 43-72. 2005.
- [2] J. Morel e G. Yu, ASIFT: A new framework for fully affine invariant image comparison. In: SIAM Journal on Image Sciences, vol. 2, issue 2, 2009.
- [3] D. Lowe, SIFT: Distinctive image features from scale invariant keypoints, International Journal on Computer Vision, vol. 60(2). pp. 91-110. 2004.
- [4] H. Y. Kim, FORAPRO: Rotation-Discriminating Template Matching Based on Fourier Coefficients of Radial Projections with Robustness to Scaling and Partial Occlusion. Pattern Recognition. vol. 43 (3), pp. 859-872. 2010.
- [5] P. Doubek, et al., Image Matching and Retrieval by Repetitive Patterns. ICPR. pp. 3195-3198. 2010.
- [6] H. Y. Kim, e G. A. Pérez, Reconhecimento de Padrões Invariante Afim Baseado nos Coeficientes de Fourier de Projeções Radiais. VI Workshop de Visão Computacional, UNESP Campus de Presidente Prudente – Brasil, 2010.
- [7] J. Morel e G. Yu, SIFT and ASIFT: IPOL Workshop 2009. <http://www.cmap.polytechnique.fr/~yu/research/ASIFT/>. Acesso em 15 dez. 2010.
- [8] K. Mikolajczyk, Harris-affine e Hessian-affine. <http://www.robots.ox.ac.uk/~vgg/research/affine/>. Acesso em 13 nov 2010.

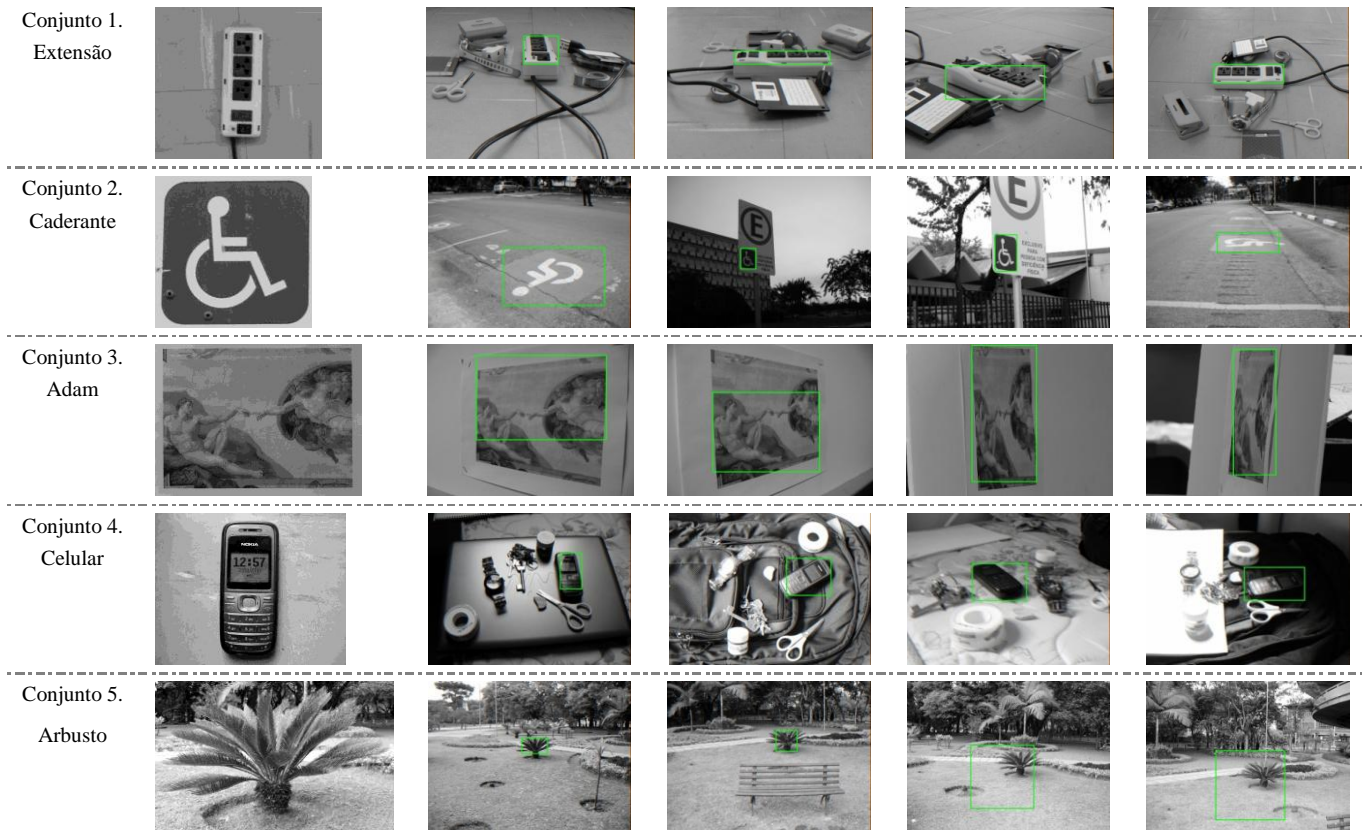


Fig. 4. Alguns exemplos dos resultados obtidos pelo Aforapro.