Instituto de Computação
Universidade Estadual de Campinas

Análise de descritores locais de imagens no contexto de detecção de semi-réplicas

Lucas Moutinho Bueno

Este exemplar corresponde à redação final da Dissertação devidamente corrigida e defendida por Lucas Moutinho Bueno e aprovada pela Banca Examinadora.

Campinas, 15 de julho de 2011.

Ricardo da Silva Torres (Orientador)

Dissertação apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

FICHA CATALOGRÁFICA ELABORADA POR
MARIA FABIANA BEZERRA MÜLLER - CRB8/6162
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA - UNICAMP

B862a

Bueno, Lucas Moutinho, 1986-
Análise de descritores locais de imagens no contexto de detecção de semi-réplicas / Lucas Moutinho Bueno. - Campinas, SP : [s.n.], 2011.

Orientador: Ricardo da Silva Torres.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Processamento de imagens. 2. Recuperação da informação. 3. Descritores. 4. Teoria bayesiana de decisão estatística. I.Torres, Ricardo da Silva, 1977-. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Informações para Biblioteca Digital

Título em inglês: Analysis of local image descriptors in the context of near-duplicate detection
Palavras-chave em inglês:
Image processing
Information retrieval
Descriptors
Bayesian statistical decision theory
Área de concentração: Ciência da Computação
Titulação: Mestre em Ciência da Computação
Banca examinadora:
Ricardo da Silva Torres [Orientador]
Humberto Luiz Razente
Cid Carvalho de Souza
Data da defesa: 19-08-2011
Programa de Pós-Graduação: Ciência da Computação

TERMO DE APROVAÇÃO

Dissertação Defendida e Aprovada em 19 de agosto de 2011, pela Banca examinadora composta pelos Professores Doutores:

Prof. Dr. Humberto Luiz Razente
CMCC / UFABC

Prof. Dr. Cid Carvalho de Souza
IC / UNICAMP

Prof. Dr. Ricardo da Silva Torres
IC / UNICAMP

Instituto de Computação
Universidade Estadual de Campinas

Análise de descritores locais de imagens no contexto de detecção de semi-réplicas

Lucas Moutinho Bueno[1]

Outubro de 2011

Banca Examinadora:

- Ricardo da Silva Torres (Orientador)
- Prof. Dr. Humberto Luiz Razente
  Universidade Federal do ABC (UFABC)
- Prof. Dr. Cid Carvalho de Souza
  Universidade Estadual de Campinas (UNICAMP)
- Prof. Dr. Eduardo Valle (suplente)
  Universidade Estadual de Campinas (UNICAMP)
- Profa. Dra. Maria Camila Nardini Barioni (suplente)
  Universidade Federal do ABC (UFABC)

[1]Suporte financeiro de CNPq e FAPESP (processo 2009/12826-5).

Resumo

Descritores locais de imagens são amplamente utilizados em diversas aplicações de reconhecimento de objetos ou de cenas. Muitos descritores locais foram propostos na literatura para caracterizar pontos de interesse em imagens. Entre eles destacam-se: PCA-SIFT, SIFT, GLOH, SURF, DAISY. Pontos de interesse em imagens são determinados por detectores. Exemplos de detectores são Harris-Affine, Hessian-Affine, Fast Hessian, MSER, DoG. O objetivo deste trabalho é investigar o uso de descritores locais no contexto de recuperação de imagens semi-réplicas por conteúdo, usando centenas de milhares de imagens. Recuperação de imagens por conteúdo consiste em achar imagens na base de dados usando o conteúdo de outra imagem como consulta, normalmente usando descritores. Imagens semi-réplicas são determinadas pela deformação de uma imagem original a partir de transformações geométricas, radiométricas ou oclusões. Devido ao grande número de pontos de interesse calculados sobre cada uma das centenas de milhares de imagens da base de dados, técnicas exaustivas de busca não são viáveis em larga escala. Assim, métodos, tais como Multicurves, LSH e Min-Hash, foram criados para melhorar a velocidade de recuperação de imagens semi-réplicas. Esse trabalho contribui para o estado da arte em dois aspectos principais. Primeiro, uma análise de descritores locais é realizada de modo a avaliar escalabilidade deles. Segundo, um sistema inovador por busca Bayesiana é proposto para diminuir significantemente a quantidade de pontos de interesse usados na recuperação de imagens semi-réplicas, sem perda significativa de acurácia.

Abstract

Local image descriptors are widely used in various applications for recognition of objects or scenes. Many local descriptors have been proposed in the literature to characterize points of interest in images. Among them are: PCA-SIFT, SIFT, GLOH, SURF, DAISY. Points of interest in images are determined by the detectors. Examples of detectors are Harris-Affine, Hessian-Affine, Fast Hessian, MSER, DoG. The objective of this work is to investigate the use of local descriptors in the context of content-based near-duplicate image retrieval, using hundreds of thousands of images. Content-based image retrieval aims at finding images in the database using the content of another image as a query, typically using descriptors. Near-duplicate images are determined by the deformation of an original image from geometric or radiometric transformations or occlusions. Due to the large number of points of interest computed on each of the hundreds of thousands images from database, exhaustive search techniques are not feasible on a large scale. Thus, methods such as Multicurves, LSH and Min-Hash, are designed to improve the speed of near-duplicate image retrieval. This work contributes to the state of the art in two major aspects. First, an analysis of local descriptors is carried out to evaluate the scalability of them. Second, an innovative system using Bayesian search is proposed to significantly decrease the amount of points of interest used in near-duplicate image retrieval, without significant loss of accuracy.

Agradecimentos

Eu gostaria de agradecer:

- aos membros da banca, pela presença e sugestões em minha defesa de mestrado;
- ao professor Ricardo da Silva Torres e a Eduardo Valle, pela orientação do trabalho;
- ao Instituto de Computação (IC) da Unicamp, e particularmente ao laboratório RECOD, pela infraestrutura provida;
- à FAPESP e ao CNPq, pelas bolsas, além de FAEPEX, CAPES e Microsoft, pelo financiamento indireto;
- a demais professores e colegas de pós-graduação, pelo apoio acadêmico, em especial a George Teodoro e Fernando Akune, por fornecimento de dados;
- à família e amigos, pelo apoio pessoal.

Sumário

Lista de Tabelas

Lista de Figuras

Capítulo 1

Introdução

Bases de imagens digitais podem ser encontradas em diversos locais, dentre os quais a internet é a mais evidente. Quando se deseja encontrar uma imagem em uma base, recorre-se a sistemas de recuperação de imagens. Esses sistemas são classicamente implementados por busca textual, em que metadados, normalmente representados por palavras-chave, são associados às imagens. A busca textual, entretanto, possui dois problemas [27]: as imagens da base devem ser previamente anotadas, algumas vezes de forma exaustiva; e o conteúdo visual (por exemplo, informações de cor e textura) das imagens não é avaliado. A recuperação de imagens por texto é dependente de interpretação humana, o que implica que uma palavra usada na busca pode não estar anotada na imagem desejada.

Em contrapartida, uma área crescente da computação, denominada recuperação de imagens por conteúdo, em inglês *Content-Based Image Retrieval* (CBIR) [6,27], não possui os problemas de recuperação de imagens por texto. Ela consiste em achar uma imagem em uma base de dados usando o conteúdo visual de outra imagem como consulta, sem depender de metadados textuais. Entende-se conteúdo visual como cor, contorno de objetos, padrões de textura ou objetos de interesse.

Uma área mais específica, mas ainda muito abrangente de CBIR, é de recuperação de imagens semi-réplicas por conteúdo, em inglês *Near-Duplicate Image Retrieval* (NDIR), também conhecida simplesmente por detecção de semi-réplicas ou detecção de semi-duplicatas. Um conjunto de pelo menos duas semi-réplicas é composto por imagens de um mesmo objeto ou cena, mas vistas de diversas formas, por perspectivas, ângulos, iluminação, resolução, etc., diferentes. Sendo assim, em um sistema de detecção de semi-réplicas deseja-se enontrar uma imagem em uma base de dados a partir de uma imagem de consulta, tal que a consulta seja uma deformação da imagem procurada. Na Figura 1.1 há três exemplos de conjuntos de semi-réplicas, formados por pares de imagens. Cada par possui uma imagem original e uma imagem transformada (deformada) da primeira. Vários exemplos de aplicações baseadas em detecção de semi-réplicas podem ser citadas.

Dentre elas, destacam-se detecção de digitais [33], de violação de direitos autorais [13], eliminação de duplicatas em bases de dados [43], identificação de obras de arte em instituições culturais [2, 40], etc.

Sistemas de CBIR são baseados em descritores, algoritmos que descrevem o conteúdo de imagens e as representam por vetores de características. Esses vetores podem ser comparados e ranqueados por uma medida (função) de distância. Descritores são divididos em dois tipos: global ou local. Descritores globais descrevem uma imagem inteira por um único vetor de características, enquanto descritores locais descrevem regiões ao redor de pontos de interesse, gerando vários vetores de características para uma mesma imagem. Pontos de interesse podem ser cantos ou *blobs* (regiões homogêneas) na imagem e são computados por detectores. Enquanto descritores globais representam melhor conceitos da visão humana, como cor, forma ou textura, descritores locais são mais robustos a transformações de imagens [19], sendo assim mais adequados para serem usados em semi-réplicas.

Idealmente, imagens diferentes de uma mesma cena ou objeto, ou seja, semi-réplicas, possuem uma mesma quantidade de pontos de interesse, nos mesmos lugares físicos do objeto ou cena. Ainda idealmente, vetores de características que representam um mesmo lugar físico são iguais, se correspondem perfeitamente e, por consequência, detectam se suas respectivas imagens são semi-réplicas. Entretanto, essa situação ideal não existe na prática, o que torna o problema mais difícil.

Imagens semi-réplicas normalmente sofreram transformações pesadas. Isso faz com que lugares físicos não estejam igualmente representados no mesmo conjunto de semi-réplicas, tanto por imprecisão na detecção de pontos de interesse, como por imprecisão no cáclulo de vetores de características. Ainda assim, descritores locais são de grande qualidade e amplamente empregados em diversas aplicações, incluindo detecção de semi-réplicas, devido à alta invariância a transformações e distinguibilidade que eles possuem [25].

Este trabalho tem por objetivo estudar descritores locais aplicados a grandes bases de dados, em específico detecção de semi-réplicas.

1.1 Motivação

A Sociedade Brasileira de Computação (SBC) definiu em [23] os grandes desafios de pesquisa em computação de 2006 a 2016. O segundo desafio diz respeito à gestão da informação em grandes volumes de dados multimídia distribuídos. A internet, cada vez mais crescente, possui uma grande quantidade de imagens digitais. A maioria dos atuais sistemas de busca por imagem encontrados na rede são textuais, e portanto dependem de metadados associados às imagens e ignora o conteúdo das mesmas. Também podem-se encontrar bases grandes de imagens além da internet, como em bancos de impressões

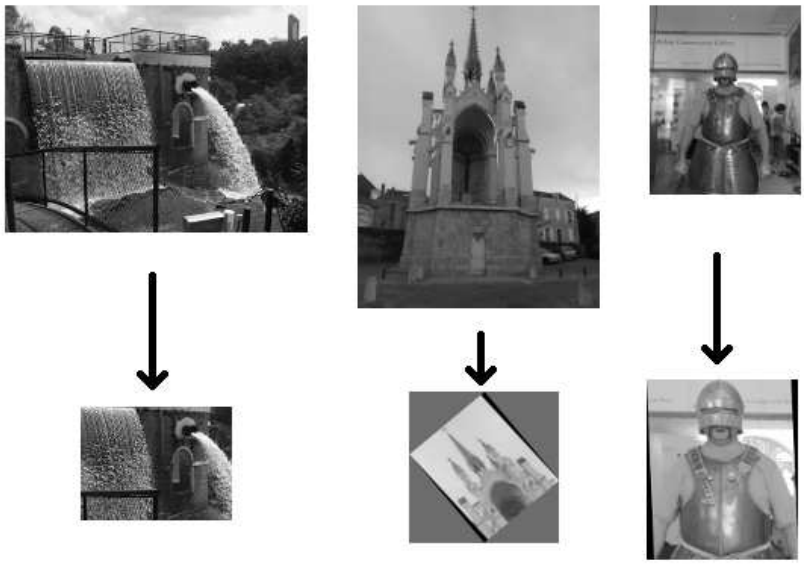

Figura 1.1: Exemplos de pares de imagens semi-réplicas, formados por imagens fotografadas (linha de cima) e transformadas (linha de baixo).

digitais, imagens médicas, galerias de arte, etc. Em alguns casos, como quando se deseja encontrar o dono de uma impressão digital desconhecida, recuperação de imagens por texto torna-se impossível. É necessário, nesses casos, avaliar o conteúdo das imagens.

Tendo esse contexto em mente, pesquisas na área de CBIR, em especial detecção de semi-réplicas, crescem continuamente. A importância do tema é testemunhada pelo grande número de publicações em diversas conferências e periódicos, tanto na área de recuperação de informação, como nas áreas de processamento de imagens e visão computacional.

1.2 Desafios de pesquisa e contribuições

Quase todos sistemas de recuperação de imagens semi-réplicas usam descritores locais e, de uma maneira ou de outra, uma função de distância para comparação: seja entre vetores de características, por algoritmos de votação; seja entre grupos de vetores, pela construção de dicionários visuais. Sendo assim duas frentes de pesquisa, distintas, porém fortemente correlacionadas, são construídas: uma apresenta contribuições científicas para o estudo de descritores locais em larga escala, outra para métodos de recuperação de imagens semi-réplicas. Para esse fim, tanto a literatura de descritores locais como a de recuperação de imagens semi-réplicas teve que ser estudada.

Apesar de alguns estudos de descritores terem sido feitos [18,25], este trabalho manteve-

se inovador em dois pontos:

- Em [25], diversos descritores são avaliados, mas aplicados somente entre pares de imagens de cenas iguais. A escalabilidade dos descritores é totalmente desconsiderada, estando fora de um contexto de detecção de semi-réplicas. Aqui, descritores são analisados em larga escala, usando centenas de milhares de imagens, aplicando-se a detecção de semi-réplicas.

- Em [18], descritores são avaliados no contexto de detecção de semi-réplicas. Entretanto, o critério de avaliação usado não considera transformações entre imagens semi-réplicas. Aqui, a qualidade de descritores é analisada em termos de invariância a transformações e distinguibilidade deles, além de explicações para escalabilidade de descritores locais serem feitas em mais profundidade.

A contribuição para métodos de recuperação de imagens semi-réplicas inclui a proposição de um novo método baseado na teoria de decisão Bayesiana. Nele, o estudo de descritores locais em larga escala é re-aproveitado e a análise do comportamento de um desses descritores (o SIFT [21]) é continuada. Sem isso, a principal vantagem do método proposto, que é a redução de consultas necessárias à base de dados para recuperar uma imagem semi-réplica, seria dificilmente alcançada. Nesse ponto, o trabalho também é inovador, pois a maioria dos métodos atuais de detecção de semi-réplicas desconsidera o comportamento do descritor usado.

As principais contribuições do trabalho são:

1. Estudo da influência do tamanho da base de imagens na qualidade de descritores locais, no contexto de detecção de semi-réplicas. Em outras palavras, estudo da escalabilidade de descritores locais, especificamente do SIFT [21] e do SURF [2], dois descritores largamente utilizados da literatura.

2. Estudo da influência da função de distância, usada para fazer correspondências entre vetores de características, na qualidade do descritor.

3. Análise refinada do comportamento das distâncias entre vetores de características, por meio de histogramas de distância.

4. Proposição de um novo método eficiente e eficaz de detecção de semi-réplicas, baseado na teoria de decisão Bayesiana [12] e na modelagem estatística dos histogramas de distância, que diminui a quantidade de consultas à base de dados para recuperar uma imagem semi-réplica.

5. Expansão do método proposto para identificar imagens sem semi-réplicas na base de dados.

6. Integração do método proposto com estruturas de indexação para melhorar ainda mais sua eficiência.

Portanto, esse trabalho apresenta contribuições científicas tanto no estudo de descritores locais, voltado à escalabilidade deles, como no desenvolvimento de um método para recuperação de imagens semi-réplicas.

1.3 Organização da dissertação

A dissertação é dividida em cinco capítulos, este primeiro e mais quatro a seguir.

No Capítulo 2, encontra-se uma revisão bibliográfica sobre descritores locais de imagem e sobre técnicas para detecção de semi-réplicas usando descritores locais.

O Capítulo 3 faz uma análise de descritores locais, especificamente do SIFT [21] e do SURF [2], no contexto de detecção de semi-réplicas. Essa análise abrange as contribuições 1, 2 e 3 listadas na seção 1.2.

O Capítulo 4 apresenta um novo sistema eficiente e eficaz de detecção de semi-réplicas baseado na teoria de decisão Bayesiana (contribuições 4, 5, 6 da seção 1.2).

Finalmente, o Capítulo 5 conclui o trabalho, retomando as contribuições, incluindo resultados obtidos, e sugerindo trabalhos futuros na área.

Capítulo 2

Conceitos fundamentais e estado da arte

Esse capítulo tem o objetivo de explicar conceitos importantes para entendimento do trabalho desenvolvido e resumir o estado da arte, tanto na área de descritores locais, como na área de detecção de semi-réplicas. Das diversas citações deste capítulo, as seguintes possuem especial relevância para o trabalho e para seu melhor entendimento [2,21,25,40]. Noções prévias de processamento e análise de imagens também ajudam o leitor. Diversos livros podem ser encontrados na literatura. Um dos mais conhecidos é [10].

2.1 Descritores Locais de Imagem

Descritores locais de imagem, computados em regiões ao redor de pontos de interesse, são usados de maneira eficaz em diversas aplicações de visão computacional e processamento de imagens, tais como: reconhecimento de objetos, reconhecimento de texturas, recuperação de imagens por conteúdo, mineração de dados de vídeo, construção de panoramas, reconstrução 3D, calibração de câmera, recuperação de imagens semi-réplicas, entre outros.

Pontos de interesse, na maioria das vezes, são vértices de contornos ou *blobs* (regiões homogêneas) na imagem. Eles são encontrados por *detectores* de pontos e possuem normalmente as seguintes informações: uma coordenada 2D na imagem, uma orientação e uma escala.

Um bom detector é determinado principalmente por sua alta covariância a transformações geométricas (como rotação e mudança de escala) e invariância a transformações radiométricas (como ruído ou variação luminosa) na imagem. Dessa forma, os mesmos pontos de interesse devem ser encontrados antes e depois de uma imagem ser distorcida, de acordo com transformação geométrica que a imagem sofreu. Um ponto de interesse

devidamente detectado antes e depois de uma transformação na imagem é *repetido*. A fração entre os pontos repetidos sobre o total de pontos encontrados é chamada de *repetibilidade*. Quanto maior a repetibilidade de um detector, melhor. Cerca de centenas a milhares de pontos de interesse são computados por imagem. Exemplos de detectores são: baseado em borda [35], baseado em intensidade [36], Harris [30], Hessian [24], Fast Hessian [2], MSER [22] e DoG [21].

Encontrados os pontos de interesse das imagens, são calculados os *vetores de características* para representarem, cada um, as regiões definidas pelos pontos de interesse encontrados (de acordo com as respectivas coordenadas, orientações e escalas dos pontos). Um vetor de característica é composto por números em R^n, em que n é sua dimensionalidade. O cálculo dos vetores de características é feito por *descritores locais*. Um bom descritor deve ser invariante a transformações na imagem e gerar vetores altamente distinguíveis. A distinguibilidade de vetores pode ser medida por uma *função de distância*, como por exemplo, as funções Euclidiana (também conhecida como L2) e de quarteirão (também conhecida como L1 ou Manhattan). Quanto maior a distância entre dois vetores mais distintos eles são entre si. Dessa forma, espera-se que pontos de interesse repetidos tenham distância mínima entre eles, isto é, sejam *invariantes a transformações*, enquanto quaisquer outros pares de pontos tenham distância maior, isto é, sejam *distinguíveis*. Exemplos de descritores locais são: shape context [3], spin images [17], complex filters [29], Steerable filters [9], PCA-SIFT [14], SIFT [21], GLOH [25], SURF [2], DAISY [34, 42] e Structural Context [20].

De maneira geral, aplicações que usam descritores locais seguem as seguintes etapas:

1. Dada uma imagem, encontram-se seus pontos de interesse.

2. Calculam-se os vetores de características para as regiões definidas pelos pontos de interesse.

3. Repetem-se os processos 1 e 2 para outras imagens diferentes da mesma cena ou objeto.

4. Dado um par de imagens, encontra-se para cada vetor de característica de uma imagem uma correspondência na outra imagem. Isso é feito a partir de uma busca por vizinho mais próximo, de forma que a distância calculada entre pares de vetores seja mínima.

5. Outros critérios podem ser aplicados para evitar correspondências espúrias, tais como consistência geométrica [41] ou limiarização das distâncias [21].

No caso específico de detecção de semi-réplicas, a busca por vizinho mais próximo é, classicamente, feita entre os vetores de uma imagem de consulta e os vetores de todas

imagens de uma base de dados. Isso é um processo custoso e exige formas alternativas, sub-ótimas, para contornar o problema (mais detalhes na seção 2.2).

Na Figura 2.1 tem-se um exemplo de uma imagem com seus pontos de interesse detectados, ilustrados por elipses (Figura 2.1a), e de correspondências entre pares de imagens de uma mesma cena, ilustrados por segmentos que unem as coordenadas dos pontos de interesse (Figura 2.1b).

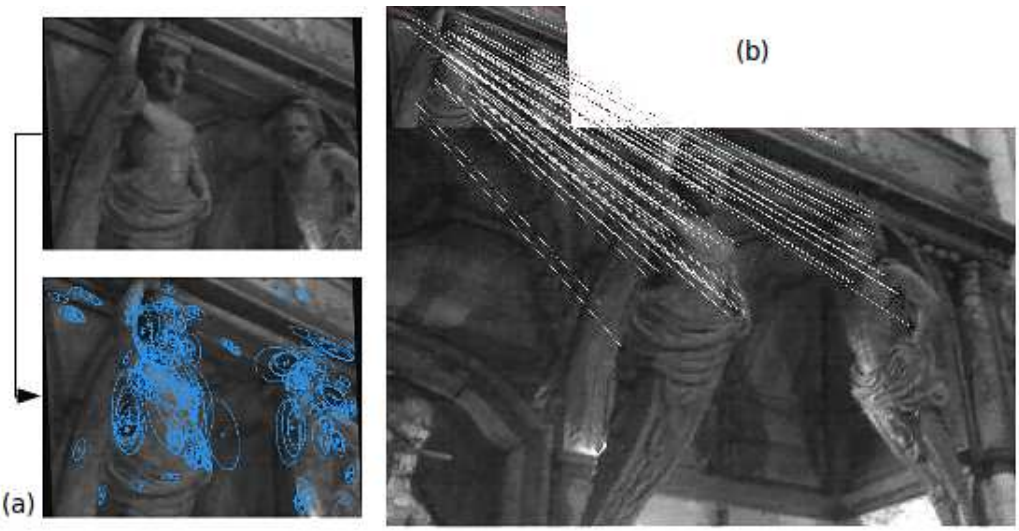

Figura 2.1: Imagem com seus pontos de interesse detectados (a) e correspondências encontradas para um par de imagens (b).

Os descritores locais mais usados podem ser divididos em dois grupos: baseados em fase e baseados em gradientes. Em [25], descritores baseados em gradientes obtiveram resultados melhores quando usados entre pares de imagens semi-réplicas (fora do contexto de recuperação em grandes bases de dados). A seguir, encontra-se um resumo de descritores locais baseados em gradiente mais conhecidos ou promissores da literatura, seguido da justificativa na escolha de trabalhar com dois deles: SIFT e SURF, com mais enfoque no primeiro.

2.1.1 Scale-Invariant Feature Transform (SIFT)

Um dos pioneiros dentre os descritores locais baseados em gradiente, o SIFT é aquele mais citado na literatura.

Em [21], é proposto um detector de pontos de interesse, Difference of Gaussians (DoG), e um descritor, o Scale-Invariant Feature Transform (SIFT).

O detector DoG (esquematizado na Figura 2.2a) consiste em identificar regiões homogêneas perto de picos locais de transição de intensidade luminosa. Isso é feito pela

convolução de filtros gaussianos com a imagem. São usadas diversas escalas de filtros e de imagem. As diferenças entre imagens convoluídas com filtros de escalas adjacentes são calculadas. Pontos em que essas diferenças são máximos locais são selecionados como candidatos a pontos de interesse. Em seguida, são descartados, dentre os candidatos, pontos em arestas e em regiões de pouco contraste da imagem. Por fim é determinada a escala e orientação principal da região entorno do ponto para que seja calculado o vetor de características do SIFT.

O vetor é formado calculando-se histogramas de gradientes. Divide-se a região em 4x4 sub-regiões e, para cada uma, seu histograma é computado com 8 bins de orientação. Um processo de normalização é feito para guardar cada dimensão do vetor em um byte e proporcionar invariância do descritor a mudanças de brilho. O total de dimensões é 128 (16 sub-regiões com 8 bins de orientação, cada). A Figura 2.2b mostra o processo de descrição do SIFT (cálculo de gradientes seguido de histogramas orientados). Para facilitar sua visualização, são ilustradas somente 2x2 sub-regiões.

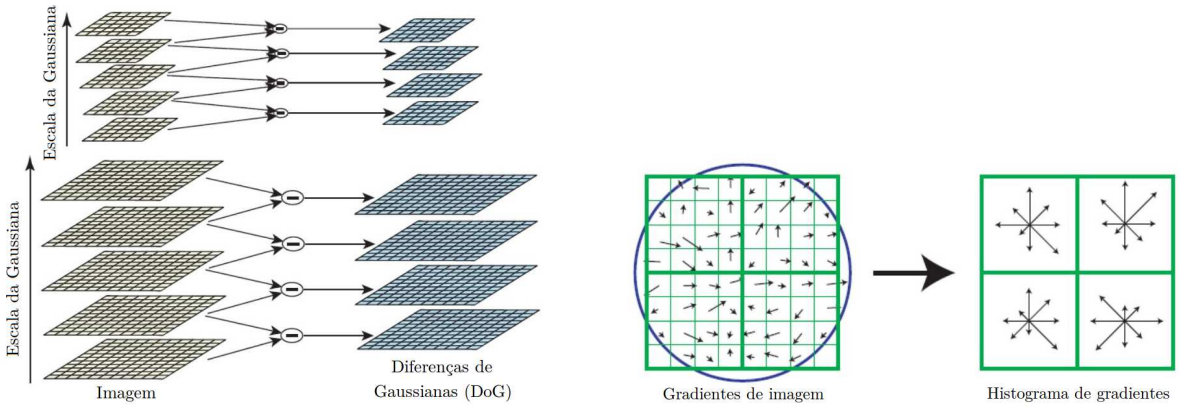

Figura 2.2: Esquematização do detector DoG (a) e do descritor SIFT (b) (traduzida de [21]).

2.1.2 PCA-SIFT

O descritor PCA-SIFT [14] contribuiu para o estado da arte ao introduzir um método redução de dimensionalidade usando uma técnica já bem conhecida, que é o *Principal Components Analysis* (PCA) [7]. Essa redução diminui espaço de armazenamento de um vetor de características e o tempo de cálculo de distâncias para encontrar correspondências. Ele mostrou ser robusto, a princípio, mas perde em precisão para o SIFT em experimentos posteriores [25].

O PCA-SIFT usa o mesmo detector de pontos de interesse do SIFT (o DoG) e o mesmo

princípio de cálculo de gradientes. Entretanto ele usa uma malha de tamanho 41x41 ao redor do ponto de interesse para o cálculo bi-direcional de gradientes, formando um vetor de 3042 dimensões. Esse vetor é reduzido a 20 componentes principais usando o método PCA. Sendo assim o vetor de características do PCA-SIFT possui apenas 20 dimensões.

2.1.3 Gradient Location and Orientation Histogram (GLOH)

Em [25], foi realizado um estudo comparativo de descritores locais em baixa escala, aplicados a correspondências entre pares de imagens distorcidas de uma cena. Foi constatado que descritores baseados em gradientes têm desempenho maior os que demais. O artigo definiu critérios de avaliação de descritores bem rígidos, usados também neste trabalho (Capítulo 3). Porém, não foi incluída detecção de semi-réplicas, cuja ordem de grandeza de dados e operações entre eles é bem maior.

Também em [25], é proposto um descritor, o Gradient Location and Orientation Histogram (GLOH), que apresenta um dos melhores resultados no próprio estudo comparativo. Ele é um descritor baseado em seus antecessores SIFT e PCA-SIFT e calcula histogramas de gradientes em uma grade circular (Figura 2.3) com 17 sub-regiões e 16 bins de orientação por sub-região, formando um vetor com 272 dimensões. O vetor de características final é reduzido para 128 dimensões usando PCA.

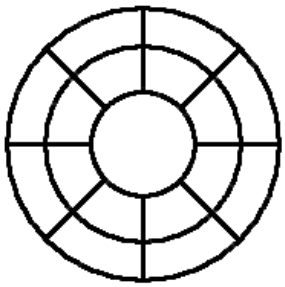

Figura 2.3: Grade circular usada para calcular os histogramas de gradientes do GLOH.

2.1.4 Speeded-Up Robust Features (SURF)

Em [2], foram propostos um novo detector de pontos de interesse (Fast Hessian) e um novo descritor: Speeded-Up Robust Features (SURF). A principal característica deles é a eficiciência, pois computam pontos de interesse e vetores de características cerca de dez vezes mais rápido que seus antecedentes e com precisão comparável.

Tanto o detector, como o descritor usam a técnica de imagem integral, em que cada pixel de uma imagem recebe um valor igual à soma dos pixeis à sua esquerda e acima,

incluindo o próprio. Isso pode ser feito em tempo linear no tamanho da imagem usando um algoritmo iterativo.

O processo de detecção de pontos de interesse usa filtros caixa (Figura 2.4b) baseados na matriz Hessiana [24], aproximando filtros gaussianos de segunda ordem (Figura 2.4a). É feita a convolução desses filtos com a imagem em diversas escalas e 4 orientações (xx, xy, yx e yy). O uso de filtros caixa sobre uma imagem integral é feito em tempo linear no tamanho da imagem, independentemente do tamanho do filtro. No caso de filtros gaussianos, a complexidade de tempo é na ordem do tamanho da imagem multiplicada pelo tamanho do filtro. Em seguida, assim como o DoG, pontos de máximo local são selecionados, eliminando-se pontos em arestas e em regiões de pouco contraste da imagem.

Já o descritor SURF divide a região entorno de um ponto de interesse em 4x4 sub-regiões. Em cada sub-região, é passado um filtro de Haar, que tem formato de caixa e, portanto, é eficiente se computado sobre uma imagem integral (Figura 2.4c), calculando a soma das respostas dos filtros e a soma do módulo das repostas dos filtros nas direções horizontal e vertical, gerando quatro valores por sub-região. Sendo assim, o vetor de características do SURF possui 64 dimensões.

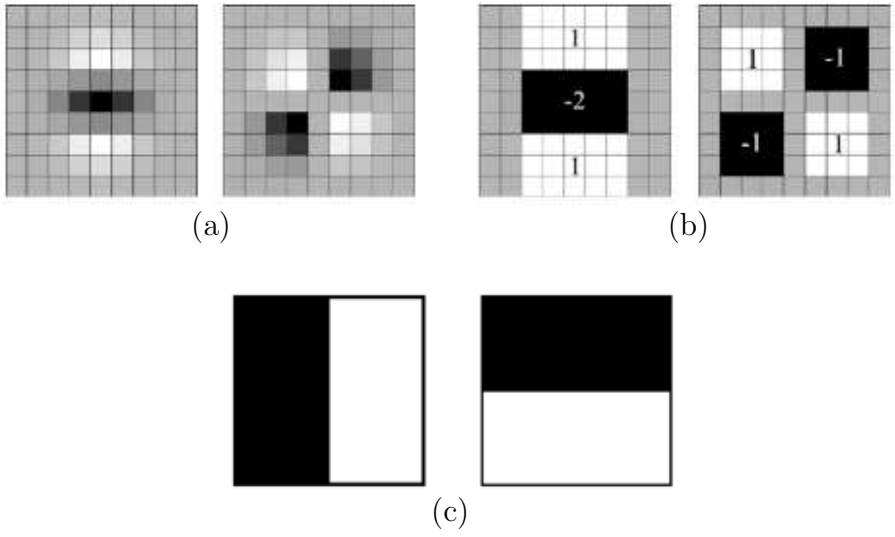

Figura 2.4: Filtros Gaussianos de segunda ordem nas direções yy e xy (a), aproximação por filtros de caixa (b), filtros de Haar (c) (reproduzidos de [2]).

2.1.5 DAISY

DAISY é um descritor, originalmente proposto em [34] e aprimorado em [42]. Ele foi desenvolvido para aplicações de reconstrução 3D de cenas a partir de imagens, com correspondência densa pixel a pixel. Isso significa que seus vetores de características são

computados para regiões entorno de cada pixel da imagem em vez de pontos de interesse. Seu desempenho é destacado principalmente por sua alta invariância a mudanças de perspectiva de câmera e presença de oclusões, entretanto não é originalmente invariante a transformações geométricas afins, como mudança de escala e rotação.

O descritor DAISY [34] calcula convoluções por filtros gaussianos sobre o gradiente de uma imagem em 8 orientações equidistantes. As convoluções são feitas em regiões dispostas em forma de anéis concêntricos em relação a um pixel central. No total são 25 regiões. Dadas as 8 orientações por região, o vetor de características do DAISY tem 200 dimensões. A Figura 2.5 ilustra os anéis concêntricos e uma das 8 orientações usadas pelo descritor.

O DAISY foi aprimorado em [42], ao ser expandido para uso em diversas aplicações. Incluiu-se nele um pré-processamento por filtros e uma redução de dimensionalidade com PCA, além dele proporcionar invariância a transformações afins. Vários parâmetros de filtros e dimensões do PCA foram testados e alguns deles superaram o SIFT para detecção de objetos, notoriamente tornando descritores baseados no DAISY promissores.

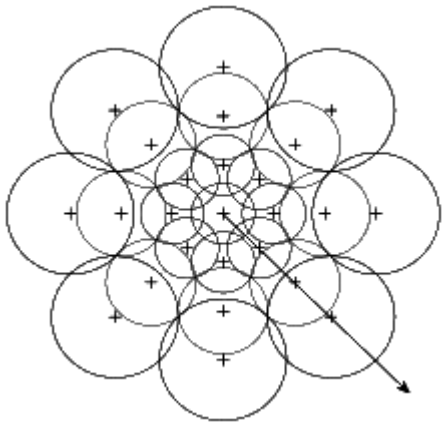

Figura 2.5: Anéis concêntricos e uma das 8 orientações usadas pelo DAISY (reproduzidos de [34]).

2.1.6 Discussão

A partir dos descritores pesquisados na literatura, escolheu-se o SIFT e o SURF, nessa ordem de prioridade, para serem usados nos experimentos de análise de escalabilidade de descritores (Capítulo 3).

O SIFT foi o principal escolhido justamente por ser um dos pioneiros na descrição local, ser o mais citado na literatura e, apesar de relativamente antigo (2004), apresentar até hoje bons resultados em publicações que o usam como base de comparação.

O SURF também foi escolhido por ser um descritor robusto a deformações, além de ser muito veloz, o que diminui o custo computacional dos experimentos. Em trabalhos recentes é, depois do SIFT, um dos descritores mais citados.

O DAISY foi descartado pois sua primeira versão não inclui invariância a todos tipos de transformações e seu aprimoramento, apesar de promissor, é protegido por direitos autorais da Microsoft Research, não estando disponível como os outros descritores.

2.2 Detecção de Semi-réplicas

Conhecendo os descritores da literatura, foquemo-nos agora em uma de suas aplicações, alvo desse trabalho de mestrado: recuperação de imagens semi-réplicas por conteúdo, ou simplesmente detecção de semi-réplicas.

Recuperação de imagens por conteúdo consiste em achar imagens na base de dados usando o conteúdo de outra imagem como consulta. Imagens semi-réplicas são geradas pela deformação de uma imagem original a partir de transformações geométricas, radiométricas ou oclusões. Portanto recuperação de imagens semi-réplicas por conteúdo consiste em encontrar imagens em um banco de dados que seja da mesma cena ou objeto de uma imagem de consulta, mas não necessariamente a mesma imagem. Na figura 1.1 podem-se ver pares de imagens semi-réplicas.

Vários exemplos de sistemas baseados em detecção de semi-réplicas ou cópia podem ser citados. Dentre eles, destacam-se detecção de digitais [33], de violação de direitos autorais [13], eliminação de duplicatas em bases de dados [43], identificação de obras de arte em instituições culturais [2, 40] , etc.

As soluções mais confiáveis para identificar corretamente uma imagem semi-réplica empregam descritores locais de uma forma ou de outra, devido à alta invariância a transformações e distinguibilidade que eles possuem, formando um sistema muito poderoso para encontrar um mesmo objeto ou cena entre diferentes imagens.

2.2.1 Sistemas baseados em votação

Uma abordagem clássica de detecção de semi-réplicas é baseada em votação, como usada em [11, 15, 38, 40]. A Figura 2.6 ilustra como é feita a recuperação de semi-réplicas por votação. É dada uma base de imagens com os pontos de interesse (*Points of interest* - PoI) previamente computados, assim como os respectivos vetores de características calculados ao redor deles (fase *offline*). Com os vetores armazenados apropriadamente, consultas por semi-réplicas já podem ser feitas (fase *online*). Dada uma imagem de consulta, repetem-se nela os processos de detecção e descrição. Em seguida, faz-se uma busca por vizinho mais próximo, ou seja, pelo vetor com menor distância, entre cada vetor da imagem de consulta

e os vetores da base. Pares de vizinhos mais próximos fazem uma correspondência, gerando um voto para a imagem correspondida na base. A imagem com mais correspondências, ou seja, a mais votada, é retornada.

Um sistema de votação pode ser estendido de maneira que cada vetor de característica da imagem consulta faça correspondência com K-vizinhos mais próximos (com $K \geq 1$), em vez de somente o primeiro. Assim, cada vetor de consulta vota K vezes, em imagens distintas ou não. Um valor de K maior que 1 pode melhorar a eficácia do sistema, porém aumentando o custo de processamento.

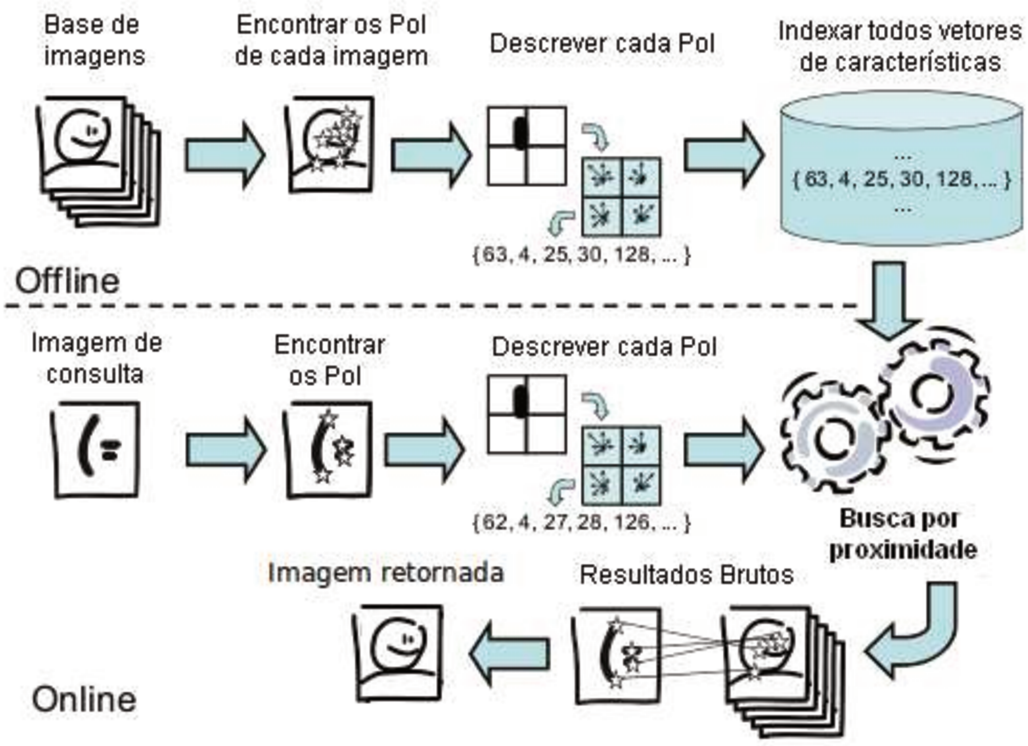

Figura 2.6: Sistema de votação para detecção de semi-réplicas (traduzido de [39]).

Por ser tolerante a falhas (apenas a maioria dos votos devem ser corretos), um sistema baseado em votação é extremamente eficaz para recuperação de semi-réplicas [40], atingindo eficácia próxima a 100%.

O grande gargalo de um sistema baseado em votação com busca exata por vizinho mais próximo, ou seja, calculando-se por força bruta as distâncias entre os vetores de consulta e todos vetores da base, é a grande quantidade de operações efetuadas para encontrar correspondências. Lembrando que cada imagem tem de centenas a milhares de pontos de interesse, uma base com 100 mil imagens tem cerca de 100 milhões de vetores

de características armazenados. A busca por vizinho mais próximo de todos os vetores de uma imagem de consulta com mil pontos de interesse faria um bilhão de comparações para retornar um resultado, o que poderia levar de vários minutos a poucas horas para recuperar uma única imagem.

Índices para busca rápida, porém aproximada, por vizinho mais próximo foram propostos para aliviar o tempo de se fazer correspondências entre vetores de características [11, 15, 38–40]. Esses índices diminuem consideravelmente o tempo de consulta para cada vetor com pouca perda de precisão. Em geral, a complexidade de tempo de busca passa de linear para logarítmica em relação ao tamanho da base (quando são implemetadas árvores [38, 39]) ou mesmo próximo de constante (quando são implemetadas tabelas hash [15]). A pouca perda de precisão da busca aproximada para cada correspondência torna-se praticamente nula quando aliada a sistemas de votação, mantendo eficácia próxima a 100%. Entretanto, ainda centenas de operações de busca devem ser realizadas, pois todos vetores da consulta são utilizados. Exemplos de índices para busca rápida são: Aproximate KNN [11], Multicurves [40], 3way-Trees [38], Locality Sensitive Hashing (LSH) [15], Floresta KD [39].

Multicurves

Uma técnica eficiente e eficaz para indexar vetores de características e possibilitar busca aproximada por K-vizinhos mais próximos é chamada Multicurves [40], usada nos experimentos do Capítulo 4. Multicurves é baseada no uso simultâneo de curvas de Hilbert [28]. Uma Curva de Hilbert é um fractal que preenche um hipercubo com qualquer número de dimensões. Cada ponto do hipercubo pode ser representado por sua projeção na curva. A Figura 2.7c esquematiza curvas de Hilbert em duas dimensões, com um ponto projetado em cada curva. Essa projeção gera um valor unidimensional, proporcional ao comprimento do caminho entre o início da curva e o ponto projetado.

Na técnica de multicurves vetores de características de dimensão N (Figura 2.7 a) são subdivididos, cada um, em c sub-vetores de dimensão N/c (Figura 2.7b) e cada um dos sub-vetores são projetados em uma curva de Hilbert (Figura 2.7c). Cada projeção gera um sub-índice unidimensional (Figura 2.7d), que é guardado em uma lista ordenada.

Na etapa de busca, cada vetor de característica é igualmente dividido em sub-vetores e projetados em curvas de Hilbert c, gerando sub-índices. Para cada sub-índice, é retornado um número fixo de candidatos a vizinho mais próximo. Os K candidados mais próximos ao vetor de consulta são retornados.

O tempo busca usando Multicurves cresce logaritmicamente em relação ao tamanho da base e a eficácia do Multicurves para detecção de semi-réplicas em um sistema de votação beira 100% [40].

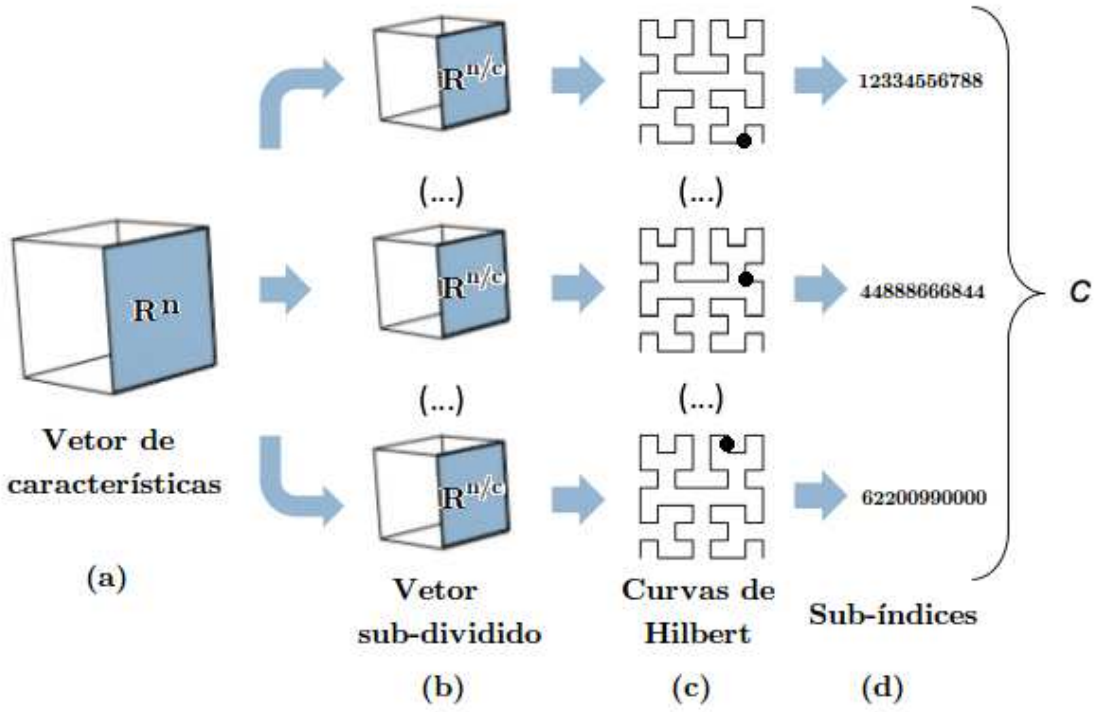

Figura 2.7: Construção de índices por Multicurves (figura reproduzida de [40]).

2.2.2 Sistemas baseados em palavras visuais

Outra abordagem popular para recuperação de semi-réplicas é a representação compacta de múltiplos vetores de características em única representação usando palavras visuais, em inglês *Bag of Words* (BoW) [31]. A Figura 2.8 ilustra como é feita a recuperação de semi-réplicas usando palavras visuais. Essa abordagem possui uma etapa de treinamento (fase offline) em que os vetores de características de uma base de imagens são agrupados usando uma técnica qualquer de agrupamento. Milhares de grupos são formados e cada grupo é representado por seu centróide, denominado uma palavra visual. O conjunto de todas palavras visuais forma um dicionário visual, que deve ser armazenado. Dessa forma imagens podem ser descritas pelas palavras visuais nelas presentes, sem necessidade de armazenamento de cada um de seus vetores de características. Uma descrição pode ser desde um vetor booleano com palavras visuais presentes a um histograma esparso de palavras. Dada uma imagem de consulta (fase online), depois dos processos de detecção e descrição, cada vetor de característica é associado à palavra visual cujo centróide possui distância mínima a ele. Assim, a imagem de consulta pode ser descrita pelas palavras visuais nela presentes e uma busca por vizinho mais próximo pode ser efetuada entre ela e as imagens da base (não mais entre vetores de características de pontos de interesse).

Funções de distância típicas para a busca são dadas pelas equações 2.1 (distância de Jaccard, no caso de vetores booleanos) e 2.2 (distância de Tanimoto, no caso de histogramas), se assemelhando a busca textual.

$$D(V1, V2) = 1 - \frac{|V1 \cap V2|}{|V1 \cup V2|} \tag{2.1}$$

$$D(V1, V2) = 1 - \frac{\sum_{i=1}^{N}(min(V1_i, V2_i))}{\sum_{i=1}^{N}(max(V1_i, V2_i))} \tag{2.2}$$

Em que D é a distância entre vetores V1 e V2; i é o índice de uma palavra visual; e N é o número de palavras visuais.

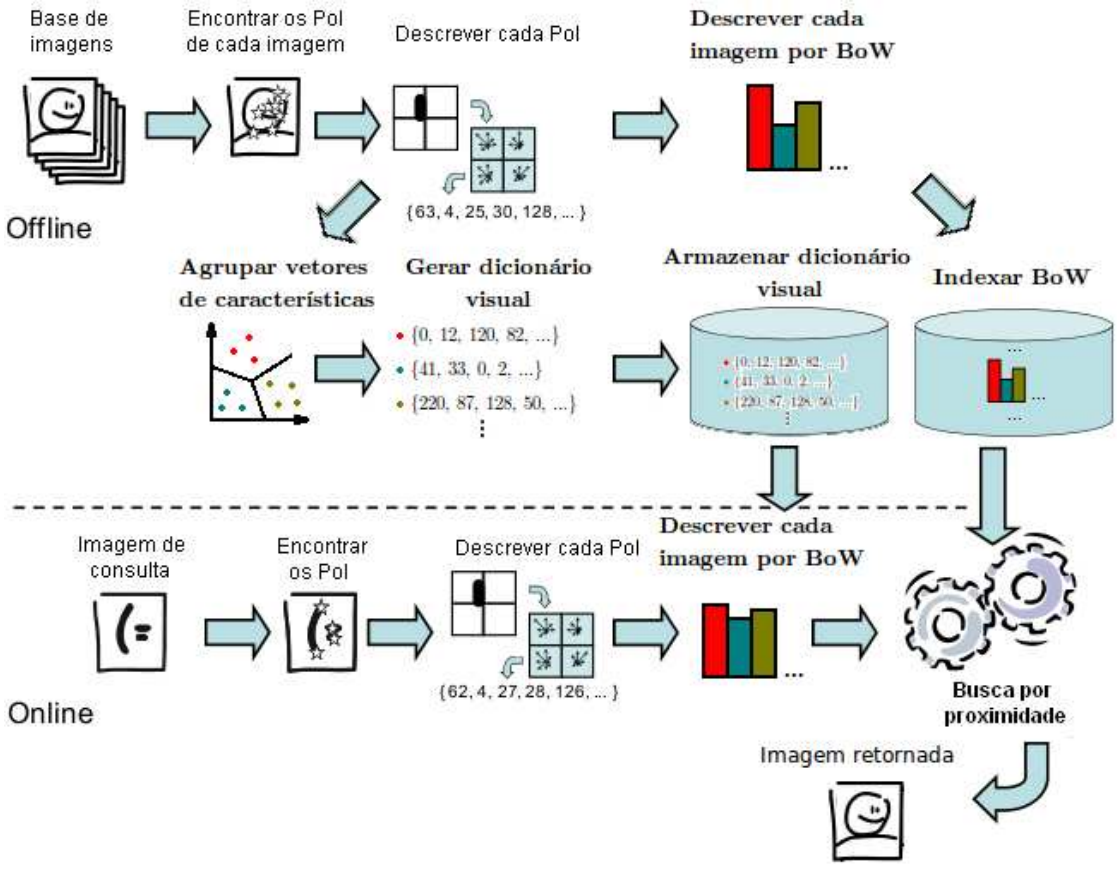

Figura 2.8: Sistema de BoW para detecção de semi-réplicas (adaptado de [39]).

Métodos que usam palavras visuais são bem rápidos, pois a recuperação de imagens é feita nos moldes da busca textual, sem usar diretamente os vetores de características das imagens de consulta. Eles também são compactos, pois não precisam de armazenar os vetores de características. Técnicas para busca aproximada também podem ser criadas

para essa abordagem [4]. Assim, métodos que usam palavras visuais tendem a ser *mais eficientes* que métodos baseados em votação.

Em contrapartida, a ausência de um sistema altamente tolerante a falhas como o de votação e a diminuição de distinguibilidade proporcionada pelo agrupamento de vetores fazem com que métodos que usam palavras visuais sejam *menos eficazes* que métodos baseados em votação.

Exemplos de métodos de recuperação de imagens que usam palavras visuais são: Video Google (para frames de vídeo) [31], min-Hash [4], baseado em entropia [44], baseado em modelo hierárquico Bayesiano [8].

2.2.3 Discussão

A grande maioria dos métodos para recuperação de semi-réplicas por conteúdo usam sistemas de votação ou de palavras visuais. Trabalhos com palavras visuais, em geral, são mais recentes, devido à crescente preocupação com eficiência ao lidar com grandes bases de dados. Entretanto, sistemas baseados em votação não são obsoletos, por causa da robustez que eles oferecem.

A grande desvantagem de sistemas de votação é a grande quantidade de vetores de características usados para se recuperar uma semi-réplica. Porém, toda essa quantidade de vetores não é realmente necessária para manter a eficácia da detecção alta. No capítulo 4, um novo método de detecção de semi-réplicas é proposto, compatível com técnicas de indexação de sistemas de votação. Entretanto, ele se difere de outros métodos da literatura por usar, em média, uma dezena de vetores de características da consulta em vez de todos eles. Em vez de votos, ele usa a regra de decisão Bayesiana para recuperar precisamente uma imagem semi-réplica com alta probabilidade de acerto (maior que 99%).

Capítulo 3

Análise de Escalabilidade de Descritores Locais

Este capítulo apresenta experimentos realizados para avaliar a escalabilidadede descritores locais, isto é, a qualidade deles em larga escala.

É importante salientar que, no caso de detecção de semi-réplicas, a quantidade de dados é bem maior que em outras aplicações que usam descritores locais, o que exige uma qualidade superior dos descritores, quanto à invariância a transformações e à distinguibilidade. Tendo isso em mente, o foco desse trabalho é avaliar os descritores propriamente ditos, deixando detectores (que têm eficácia invariante ao tamanho da base de imagens) em segundo plano.

Foram selecionados dois descritores para serem avaliados: SIFT [21] e SURF [2]. O SIFT foi escolhido por ser o descritor local mais referenciado na literatura, por ser um dos pioneiros e ser de alta qualidade [21, 25]. O SURF foi escolhido por sua extrema rapidez de computação, e ainda ser comparável ao SIFT em precisão [2].

3.1 Metodologia de Avaliação

Avaliar a qualidade de um descritor em larga escala significa saber quão capaz ele é de fazer correspondências corretas entre pares de vetores. Para isso deve-se antes definir quando que uma correspondência é correta. Primeiramente, gostaria-se que uma correspondência seja feita entre pares de vetores de imagens semi-réplicas. Esse critério já é considerado suficiente em [18], se o objetivo for somente a recuperação da imagem em si. Entretanto, pode ser que a correspondência, a princípio correta pelo critério de encontrar a imagem desejada, se refira a pontos de interesse totalmente distintos. Isso fere o princípio do descritor local de ser invariante a transformações. Em [25], uma correspondência é correta quando for feita entre vetores referententes a pontos de interesse repetidos, ou seja,

além das imagens correspondidas serem semi-réplicas, a correspondência deve obedecer a transformação geométrica entre elas, em termos de coordenadas no espaço de imagem, escala e rotação dos pontos de interesse repetidos.

Deve-se, então, estabelecer um critério objetivo para determinar se um par de pontos entre imagens semi-réplicas é repetido ou não. Assim como em [25], o critério estabelecido foi que: se a projeção da elipse, região formada em função das coordenadas, escala e rotação de um ponto de interesse de uma imagem original sobre outra elipse formada por um ponto de interesse de sua imagem semi-réplica for maior que 60%, então os pontos são repetidos. Como a transformação entre um par de imagens semi-réplicas (original e de consulta) é conhecida, a projeção da elipse definida para um ponto de interesse pode ser obtida pela matriz de transformação entre as imagens.

A Figura 3.1a ilustra um exemplo de projeção, em que a elipse formada por um ponto de interese da imagem original (ao centro) é projetada sobre um ponto de interesse da imagem de consulta (à esquerda), formando uma intercecção (à direita). Esse critério, apesar de difundido na literatura [25] se mostrou, em experimentos realizados neste trabalho, parcialmente impreciso no caso específico do DoG (detector do SIFT). Nesse caso, alguns pares de pontos de interesse de imagens de consulta se diferenciam apenas por uma rotação que os deixam em sentidos opostos, permanecendo com as mesmas escalas e coordenadas (Figura 3.1b). Diz-se que esses pares são ambíguos em termos de determinação de pontos repetidos. A projeção da elipse formada por um ponto de interesse de uma imagem original sobre esses pares de pontos ambíguos possui a mesma intersecção para ambos os pontos. Se essa intersecção for maior que 60%, ambos pontos seriam considerados repetidos. Entretanto, como eles possuem orientações opostas, apenas um vetor de características deles possui distância baixa em relação ao vetor da imagem original. O que significa que apenas um dos pontos ambíguos, aquele com sentido mais próximo à projeção do ponto da imagem original, é realmente repetido. No exemplo da Figura 3.1b, dos pontos de interesse ambíguos da imagem de consulta, apenas aquele com orientação para cima, realmente se repete na imagem original.

Outro fator que se deve levar em conta ao avaliar um descritor é descartar erros do detector. Se entre pares de imagens semi-réplicas não há um determinado ponto de interesse repetido, ou seja, que é detectado somente em uma das imagens então não é possível o vetor de características encontrar seu par apropriadamente na outra imagem. Assim sendo, define-se uma medida de avaliação de acurácia do descritor chamada *matching score*, que consiste na *razão entre a quantidade de correspondências entre vetores de pontos repetidos sobre o total de pontos repetidos nas imagens semi-réplicas.*

Para possibilitar a análise de descritores em alta escala, seguiu-se um fluxograma de atividades, apresentado na Figura 3.2. Cada atividade (etapa) é representada por um retângulo do fluxograma, numerada, para referência (entre parênteses no texto), de 1 a

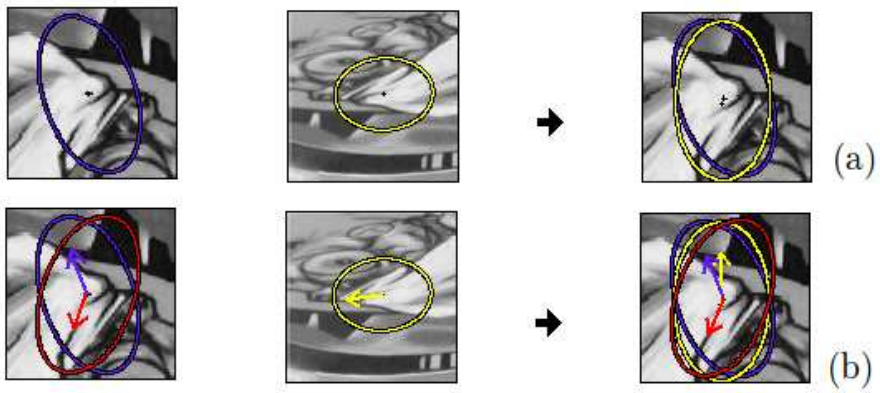

Figura 3.1: Projeção da elipse formada por um ponto de interesse de uma imagem original sobre uma semi-réplica. Sem pontos ambíguos (a) e com pontos ambíguos (b) (figura modificada de [25]).

12. As setas indicam relação de dependência entre as atividades, desde o início até o fim. Considere três bases de imagens:

    - Base de imagens de confusão (etapa 1): constituída por até centenas de milhares de imagens, serve para testar a escalabilidade do descritor em sistemas de detecção de semi-réplicas. Foi usada uma base da Yahoo disponível ao Instituto de Computação, com 110 mil imagens. Ela foi dividida em três tamanhos, que variam entre 0% (sem imagens de confusão), 50% (55 mil imagens) e 100% (110 mil imagens).

    - Base de imagens originais (etapa 2): constituída de centenas de imagens. Foram usadas 225 imagens pessoais, fotografadas ou de uma câmera de alta definição ou de um telefone celular, em nossos experimentos. Foi tomado o cuidado para que as imagens fotografadas fossem específicas o suficiente para não haver semi-réplica na base de confusão.

    - Base de imagens de consulta (etapa 3): constituída por transformações artificiais conhecidas das imagens originais. Foram usados cisalhamento, rotação, mudança de escala, ruído gaussiano e *dithering*, combinados ou individualmente, em nossos experimentos. Todas as imagens também sofreram recorte[1]. 163 das imagens originais geraram, cada uma, uma imagem de consulta. Na Figura 1.1 podem-se ver exemplos de transformações.

As imagens de cada base tiveram seus pontos de interesse detectados e descritos (etapas 4, 6 e 7) por detectores e descritores previamente implementados (etapa 5).

[1]As transformações nas imagens originais foram efetuadas usando o comando *convert* do linux. Em http://www.imagemagick.org/script/convert.php encontram-se as especificações do comando e das transformações (último acesso em 07/07/2011).

Montou-se um sistema de detecção de semi-réplicas por votação, como explicado no Capítulo 2, usando busca exata por vizinhos mais próximos. Como o foco do trabalho é avaliar a qualidade dos descritores, não foram usadas estruturas de indexação para buscas aproximadas, para não perder precisão.

Na fase "offline"do sistema, armazenaram-se juntamente os vetores de características das bases original e de confusão (etapa 10) para os diversos tamanhos da base de confusão (etapa 9).

Na fase "online", encontraram-se as correspondências exatas entre todos vetores de todas imagens de consulta e os vetores armazenados (etapa 11). As distâncias entre os vetores das imagens de consulta e os 5-vizinhos mais próximos das bases original e de confusão foram armazenadas para agilizar experimentos posteriores. O cálculo de correspondências exatas com a base de confusão é extremamente custoso em termos computacionais, e foi obtido usando um sistema de computação paralela de alta performance [32][2].

Como as imagens da base de consulta foram geradas por transformações conhecidas das imagens originais, é possível calcular as correspondências corretas (etapa 8), de acordo com o critério definido acima, que considera pontos repetidos.

Com as distâncias armazenadas e os pontos repetidos definidos, o *matching score* foi calculado (etapa 12). Para perceber o efeito causado no *matching score* pelo incremento de vizinhos mais próximos, consideraram-se tanto correspondências com vizinho mais próximo como correspondências com os 5-vizinhos mais próximos.

Finalmente, uma análise dos descritores foi feita (etapa 13), como descrito adiante.

3.2 SIFT

Essa seção mostra a análise dos resultados provenientes dos experimentos com o descritor SIFT [21] (com detector DoG) no contexto de recuperação de semi-réplicas. Todos os experimentos usaram a implementação original do SIFT, com 128 dimensões[3]. A seção é dividida em três subseções, cada uma dedicada para um tipo de avaliação do SIFT:

1. Análise de pontos repetidos ambíguos do DoG. Essa etapa do trabalho mostra que critério de determinação de pontos repetidos em [25] é insuficiente e diminui a eficácia real do descritor.

2. Análise de escalabilidade do SIFT. Essa etapa mostra que o SIFT escala bem à medida que a base de confusão cresce. Também explica alguns motivos para boa escalabilidade do descritor por meio de histogramas de distâncias.

[2]Agradecemos a George Teodoro por fazer o cálculo de correspondências entre vetores de características do descritor SIFT.

[3]http://www.cs.ubc.ca/~lowe/keypoints/ (último acesso em 07/07/2011).

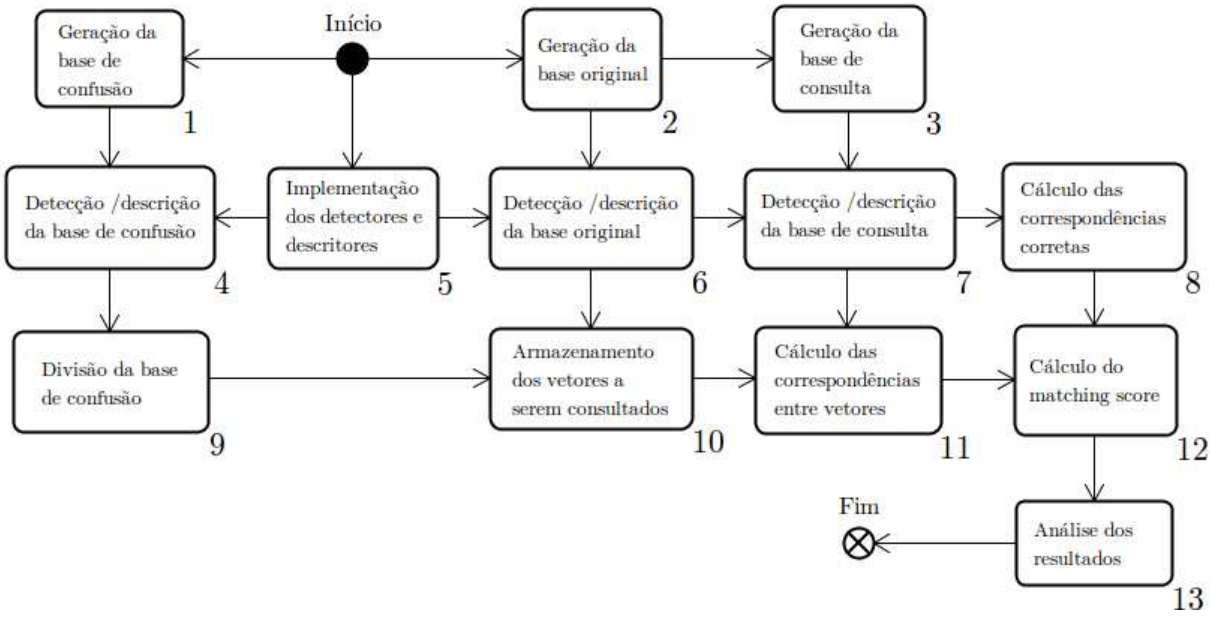

Figura 3.2: Fluxograma de atividades realizadas para avaliar descritores locais em alta escala.

3. Análise de transformações. Essa etapa mostra como o SIFT se comporta de acordo com diversas transformações em imagens, generalizando um modelo estatístico para o descritor.

3.2.1 Análise de pontos repetidos ambíguos

Como dito na Seção 3.1, o critério de determinação de pontos repetidos em [25] não diferencia o sentido de orientação dos pontos de interesse, que permanece oculto quando as regiões entorno dos pontos são representadas por uma elipse. Para ilustrar esse problema com o DoG, um histograma de distâncias entre vetores de características SIFT de pontos repetidos foi computado. Para isso, contaram-se os valores das distâncias entre todos os vetores de pontos de interesse repetidos das bases original e de consulta. Os histogramas computados foram de 50 bins de tamanho 15, cada. O tamanho dos bins foi escolhido para ser pequeno o suficiente para representar as reais distâncias entre vetores de características e grande o suficiente para evitar ruídos. A função de distância usada foi a Euclidiana.

A Figura 3.3a mostra o histograma de distâncias entre vetores de características de pontos repetidos de acordo com o critério de [25]. Pode-se notar que o histograma possui duas modas: uma maior, em torno de 80 e outra menor, em torno de 500. Considerando apenas pontos ambíguos, como explicado na Seção 3.1, o histograma gerado é aquele da Figura 3.3b. Percebe-se claramente, que metade dos pontos ambíguos pertence a uma

moda enquanto a outra metade pertence a outra. Esse fato corrobora a análise de que apenas um ponto de interesse de cada par de pontos ambíguos é verdadeiramente um ponto repetido.

De acordo com experimentos realizados, o matching score do SIFT, usando toda a base de confusão, correspondências com 5-vizinhos mais próximos dos vetores de consulta, e distância Euclidiana, sobe de 74,0% para 89,5% apenas se retirando pontos ambíguos do cálculo.

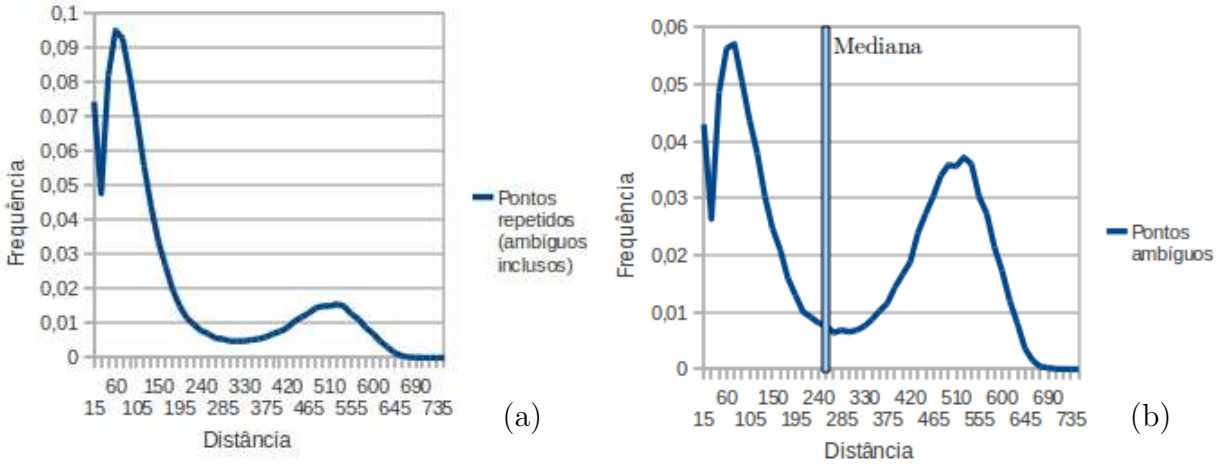

Figura 3.3: Histograma de distâncias entre vetores de características de pontos repetidos de acordo com o critério de [25] (a) e Histograma de distâncias entre vetores de pontos ambíguos (b).

3.2.2 Análise de escalabilidade

Retirando pontos ambíguos, o SIFT (com detector DoG) foi testado para diferentes tamanhos de base de confusão. O matching score encontrado, por tamanho de base, e por parâmetro K, se encontra na Tabela 3.1. Percebe-se uma característica bastante peculiar nos resultados. A acurácia do descritor cai muito lentamente com o aumento da base, permanecendo quase estável na faixa dos 90% para correspondências com 5-vizinhos mais próximos. Isso significa que o SIFT é escalável quanto ao tamanho da base e adequado para ser usado em aplicações de recuperação de semi-réplicas por conteúdo em grandes bases de imagens.

Foi usada distância Euclidiana no cálculo de correspondências. A distância de quarteirão também foi testada para o caso sem a base de confusão, porém os resultados não se alteraram significantemente.

Para saber os motivos da alta escalabilidade do descritor, as distâncias entre vetores

Tabela 3.1: Acurácia do descritor SIFT para diferentes tamanhos da base de imagens.

Número de vetores de características	1 milhão	35 milhões	70 milhões
Número de imagens	225	55 mil	110 mil
Matching score ($K = 5$)	92,0%	89,9%	89,5%
Matching score ($K = 1$)	90,0%	87,6%	87,1%

de pontos repetidos foram avaliadas. Assim, dois histogramas de distância foram computados: o primeiro conta os valores das distâncias entre todos os vetores de pontos de interesse repetidos das bases original e de consulta (como na Figura 3.3, mas sem os pontos ambíguos); o segundo conta distâncias entre vetores de pontos não repetidos tomados aleatoriamente. Um total de 1 milhão de pontos aleatórios foram tomados. Os histogramas computados foram de 50 bins de tamanho 15, cada.

Assim, o descritor SIFT só pode ser escalável se a intersecção dos histogramas computados for mínima. É o que se vê na Figura 3.4. A separabilidade é alta a ponto de nenhuma distância entre vetores de pontos não repetidos ser menor que 75 e não haver mais de cinco pares de vetores de pontos não repetidos com distância menor que 105.

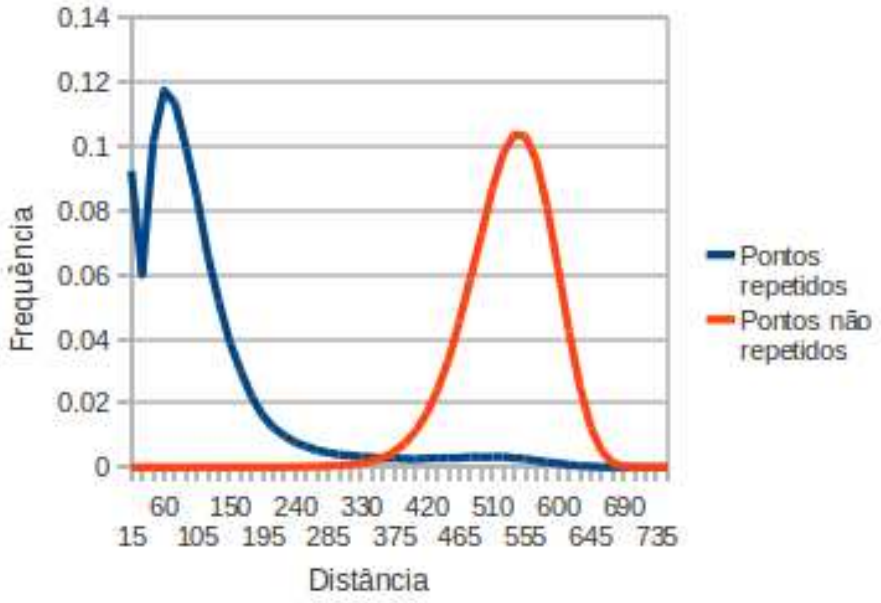

Figura 3.4: Histogramas de distâncias entre vetores de características de pontos repetidos e pontos não repetidos, para o SIFT.

Outro fator que contribui para a escalabilidade do descritor é a correlação positiva entre

as distâncias entre vetores de pontos repetidos e não repetidos. Criou-se um histograma em $\mathbb{N}^2 \rightarrow \mathbb{N}$ da seguinte forma: as distâncias entre todos os vetores de pontos de interesse repetidos das bases original e de consulta foram calculadas. Os valores dessas distâncias foram contados juntamente com os valores de distâncias entre seus respectivos vetores de consulta e vetores de pontos não repetidos tomados aleatoriamente, formando assim tuplas com duas distâncias entre vetores. Um total de 5 milhões de pontos aleatórios foram tomados. O histograma computado foi, nas duas dimensões, de 50 bins de tamanho 15.

A Figura 3.5a mostra o histograma de distâncias que correlaciona distâncias entre vetores de pontos repetidos e não repetidos. A Figura 3.5b representa o mesmo histograma em escala logarítmica, que facilita a visualização de valores baixos. Note que quando as distâncias entre vetores de pontos repetidos são maiores, as distâncias entre vetores de pontos não repetidos tendem a ser maiores também, contribuindo para que distâncias entre vetores de pontos repetidos permaneçam mínimas e a acurácia do descritor permaneça alta.

A linha diagonal que parte da origem do histograma na Figura 3.5 o divide em duas partes: a de cima representa os casos em que as distâncias entre vetores de pontos repetidos são menores que as distâncias entre vetores de pontos não repetidos; a de baixo representa os casos contrários. Para 89% do histograma, ou mais especificamente, para toda a parte em que a distância entre vetores de pontos repetidos não passa de 165, menos de 5 casos se encontram na parte de baixo do histograma. Em outras palavras, de 5 milhões de tuplas de distâncias do histograma, 89% delas (aquelas com as menores distâncias entre vetores de pontos repetidos) tem menos de 5 valores de distâncias entre vetores de pontos não repetidos menores que entre vetores de pontos repetidos. Esse resultado é compatível com aquele da Tabela 3.1, em que o matching score considerando 5-vizinhos mais próximos é próximo desse valor.

É importante salientar que, mesmo que os valores observados no histograma sejam próximos aos da tabela, a variância na cauda do histograma é muito grande para constatar uma prova empírica de implicação entre ele o matching socre. Mesmo com 5 milhões de tuplas, múltiplas instâncias do histograma podem variar o valor obtido de 89% (quando menos de 5 tuplas se encontram na parte de baixo do histograma) em até 5%. Se fosse considerado somente uma tupla em vez de 5, essa variância seria ainda maior. De qualquer maneira, o hitograma ilustra adequadamente o comportamento das distâncias entre vetores de características do SIFT e explica sua escalabilidade.

3.2.3 Análise de transformações

Nessa seção uma análise do comportamento das distâncias entre vetores de pontos repetidos do SIFT foi feita para diversas transformações de imagens. São elas: recorte, escala,

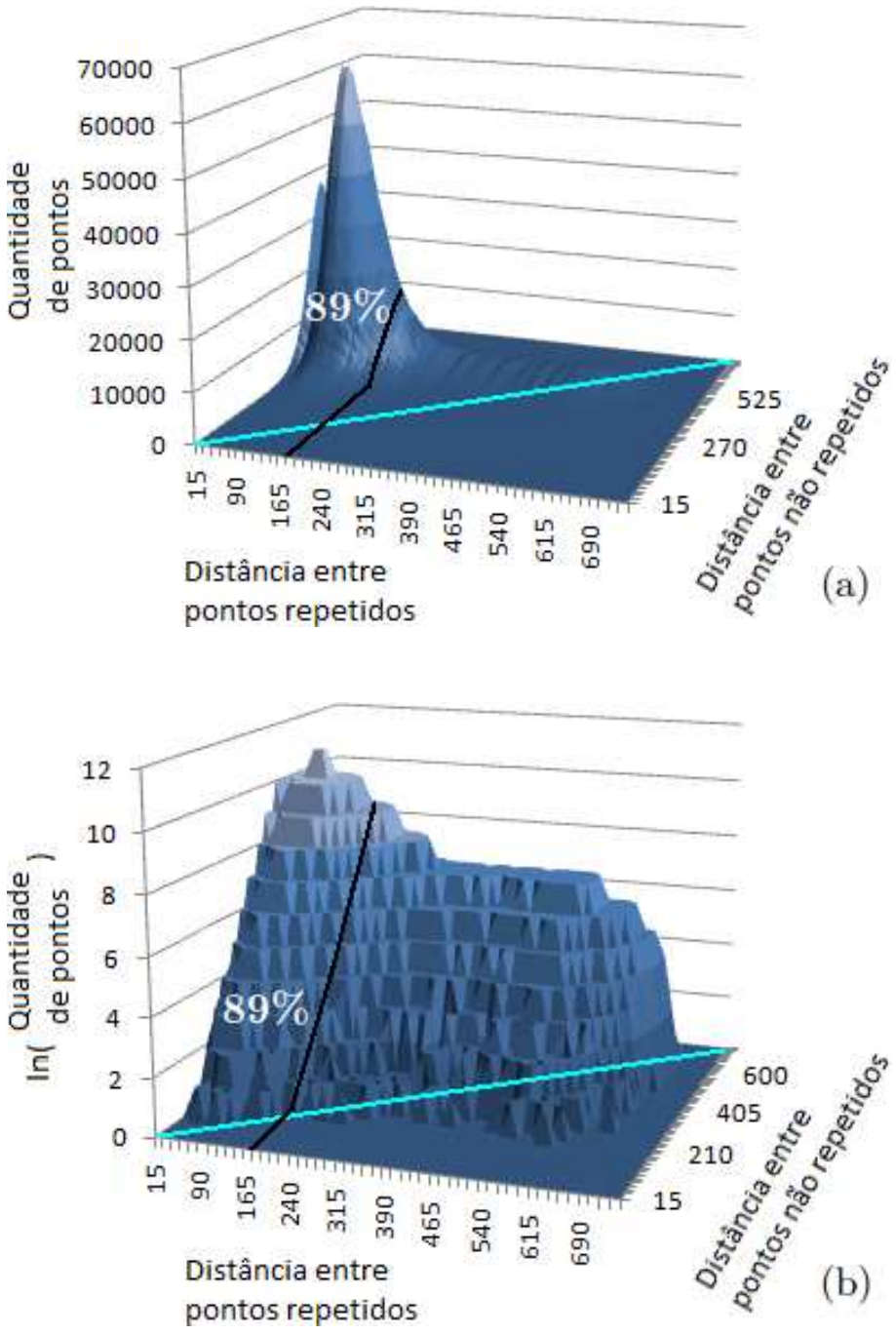

Figura 3.5: Histogramas de distâncias do SIFT correlacionando pontos repetidos e não repetidos. Em números absolutos (a) e em escala logarítmica (b).

rotação, cisalhamento, ruído gaussiano, correção gamma de luminosidade e *dithering*. A função de distância usada foi a Euclidiana.

O objetivo dessa análise é mostrar que as distâncias possuem, para cada transformação, um padrão de frequência, e que esse padrão é generalizável e modelável estatisticamente. Isso servirá de base para um novo método de recuperação de imagens semi-réplicas, descrito no Capítulo 4.

As transformações foram analisadas tanto individualmente como combinadas entre si.

Transformações combinadas Para as transformações combinadas, em que mais de um tipo de transformação pode coexistir, as imagens da base de consulta foram divididas nos seguintes grupos:

1. imagens contendo somente transformações geométricas (escala, rotação e cisalhamento), além de recorte;
2. imagens contendo somente transformações radiométricas (ruído gaussiano e *dithering*), além de recorte;
3. imagens contendo somente recorte;
4. somente imagens geradas a partir de fotos em alta resolução;
5. todas imagens.

Os três primeiros grupos acima são mutualmente exclusivos, enquanto o quarto intercepta parte dos anteriores. O quinto grupo é a união dos anteriores, somado a imagens pertencentes a nenhum outro grupo (por exemplo, imagens de baixa resolução com rotação e *dithering*).

A Figura 3.6 mostra que o primeiro e o segundo grupo possuem histogramas de distâncias entre vetores de pontos repetidos semelhantes, gerando curvas assimétricas com moda por volta de 60, o que indica um padrão de frequência para diferentes tipos de transformação. O quarto grupo praticamente coincide com o quinto, o que indica que a resolução da imagem pouco influencia no comportamento das distâncias entre vetores de pontos repetidos. O terceiro grupo é analisado mais profundamente a seguir.

Recorte O recorte é a transformação mais particular de todas. Quando ela está presente (não necessariamente sozinha), cerca de 4% das distâncias entre vetores de pontos repetidos ficam acima de 300, mais que qualquer outra transformação. Entretanto, quando recorte é a única transformação efetuada, cerca de 96% das distâncias são zero (Figura 3.6). Esse fato se deve a oclusões nas bordas das imagens recortadas. O descritor perde

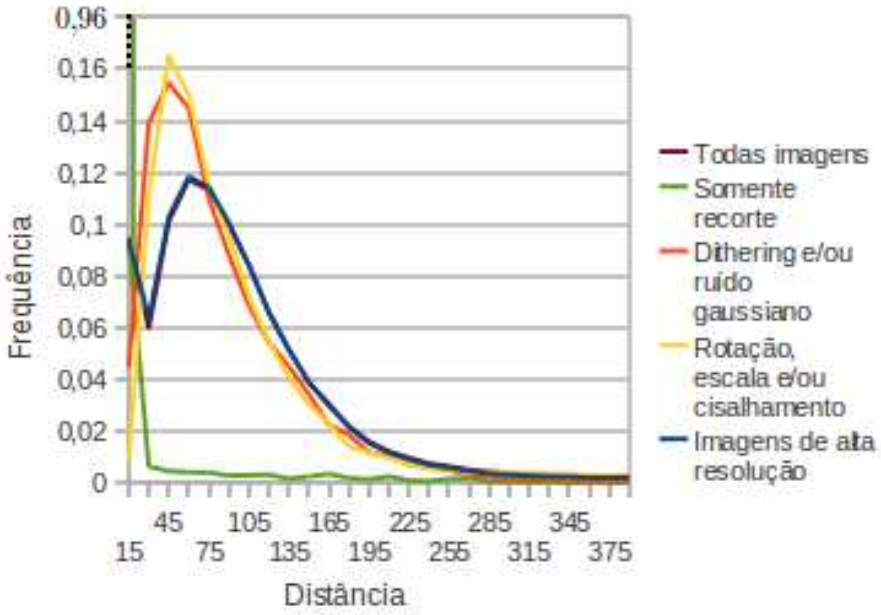

Figura 3.6: Histogramas de distâncias entre correspondências para várias transformações (transformações juntas).

significantemente sua propriedade de invariância quando descreve regiões que estão parcialmente presentes em imagens semi-réplicas.

A Figura 3.7 ilustra o comportamento do descritor em áreas de oclusão. Os retângulos desenhados nas imagens originais de cima representam os recortes efetuados na geração das imagens de baixo. Os pontos destacados nas imagens de baixo representam pontos de interesse repetidos cujas distâncias entre os vetores de características são altas, com valores maiores que 300. Como pode ser observado, as distâncias altas são mais frequentes entre vetores de pontos próximos às bordas do recorte, atingindo 100% de frequência no caso de imagens em que recorte é a única transformação efetuada.

Transformações individuais Para a análise de transformações individuais foi usada outra base de imagens originais, contendo 44 imagens com 112 mil pontos de interesse. A troca de base reforça a generalidade da análise. Além disso, essa nova base possui imagens mais diferenciadas, incluindo imagens com diversas resoluções, imagens com ou sem compressão jpeg, imagens nítidas ou embassadas, fotos não modificadas, fotos modificadas por computador ou mesmo figuras geradas por computador.

Para cada imagem original foram geradas diversas imagens semi-réplicas por transformações de cisalhamento, correção gamma de luminosidade, escala, rotação e ruído gaussiano com diversos parâmetros. A Figura 3.8 exemplifica uma série de transformações efetuadas, na ordem citada, sobre uma das imagens originais da base.

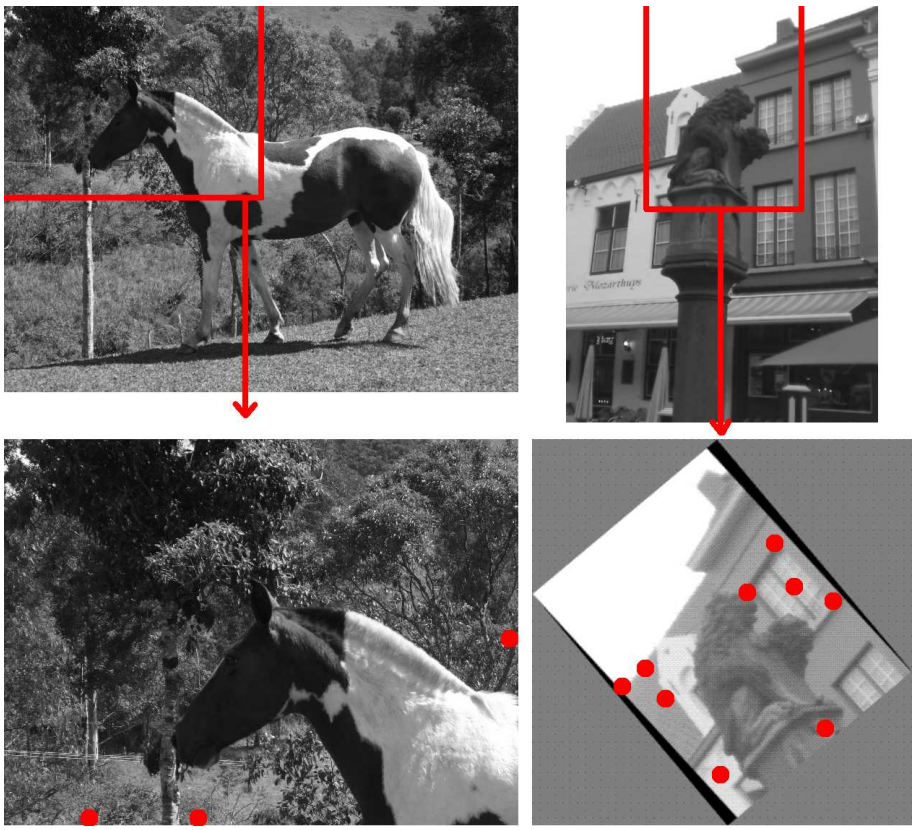

Figura 3.7: Exemplo de imagens recortadas. Pontos repetidos com distância alta em destaque.

A Figura 3.9 mostra os histogramas de distâncias entre vetores de pontos repetidos computados para essas transformações. Notam-se os seguintes padrões de distâncias para as seguintes transformações:

- escala: padrão de distribuição assimétrico das distâncias com moda entre 40 e 80, semelhante para diversos parâmetros, que variam de redução a 25% até ampliação a 200% da imagem original;
- rotação: padrão de distribuição assimétrico das distâncias com moda entre 50 e 60, semelhante para diversos parâmetros, que variam de rotação de 10^o a 180^o. Possui distâncias notoriamente inferiores no caso de rotações de 90^o e 180^o.
- cisalhamento: padrão de distribuição assimétrico das distâncias com moda crescente à medida que o parâmetro aumenta. Mostra a baixa invariância do SIFT a cisalhamentos de ângulos altos (a partir de 18^o);
- correção gamma: padrão de distribuição assimétrico das distâncias com moda entre 30 e 70, com alta frequência de distâncias zero. Mostra a invariância perfeita do SIFT a mudanças de liminosidade para cerca de 10% dos vetores de características. A perfeição para alguns vetores se dá pela forte normalização que o descritor sofre;
- ruído gaussiano: padrão de distribuição assimétrico das distâncias com moda entre 20 e 40. Possui uma segunda moda, simétrica, em torno de 100, que se mostra mais evidente à medida que o parâmetro aumenta.

Nota-se, na soma das frequências das distâncias para todas as transformações (Figura 3.9), que existe um padrão de distribuição assimétrico, que pode ser modelado estatisticamente. É importante observar também que o comportamento dessa curva se assemelha àquelas encontradas na Figura 3.6 (excluindo transformações somente de recorte). Isto evidencia que a mudança da base de imagens originais pouco afeta o comportamento das distâncias entre vetores de características do SIFT. Sendo assim, um modelo estatístico dos histogramas pode ser ajustado de forma robusta considerando apenas um conjunto suficientemente grande de imagens e transformações.

O próximo capítulo mostra que o modelo Chi se encaixa bem nos histogramas, aliando-se a um método eficiente e eficaz de recuperação de semi-réplicas baseado na teoria de decisão Bayesiana.

3.3 SURF

Essa seção mostra a análise dos resultados provenientes dos experimentos com descritor SURF [2] (com detector Fast Hessian) [2] no contexto de recuperação de semi-réplicas.

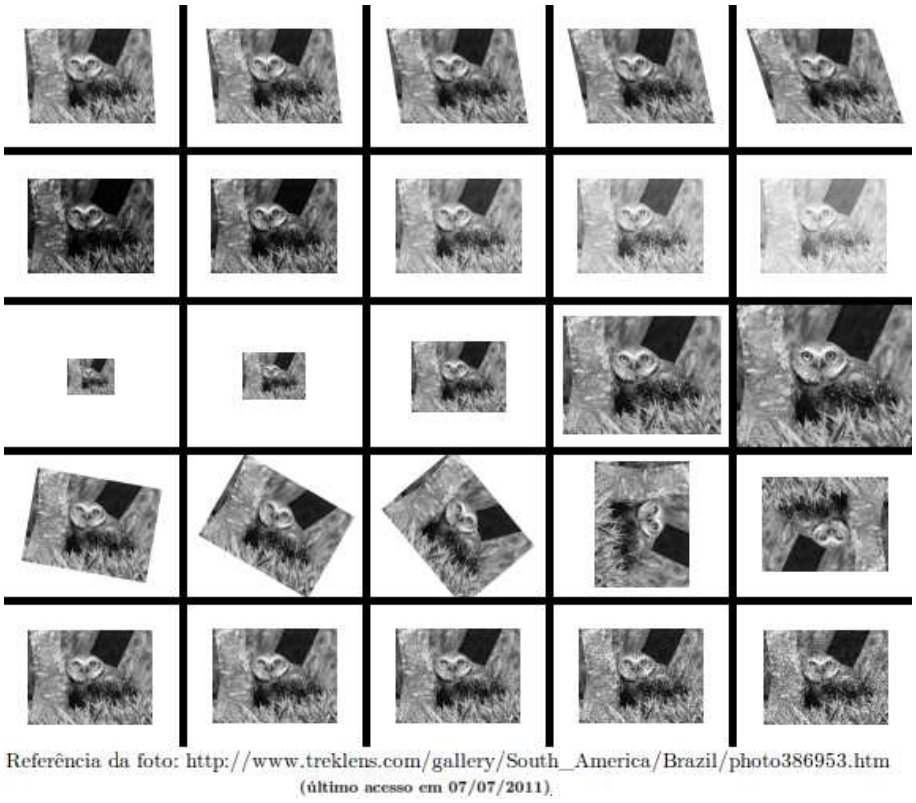

Figura 3.8: Exemplos de transformações.

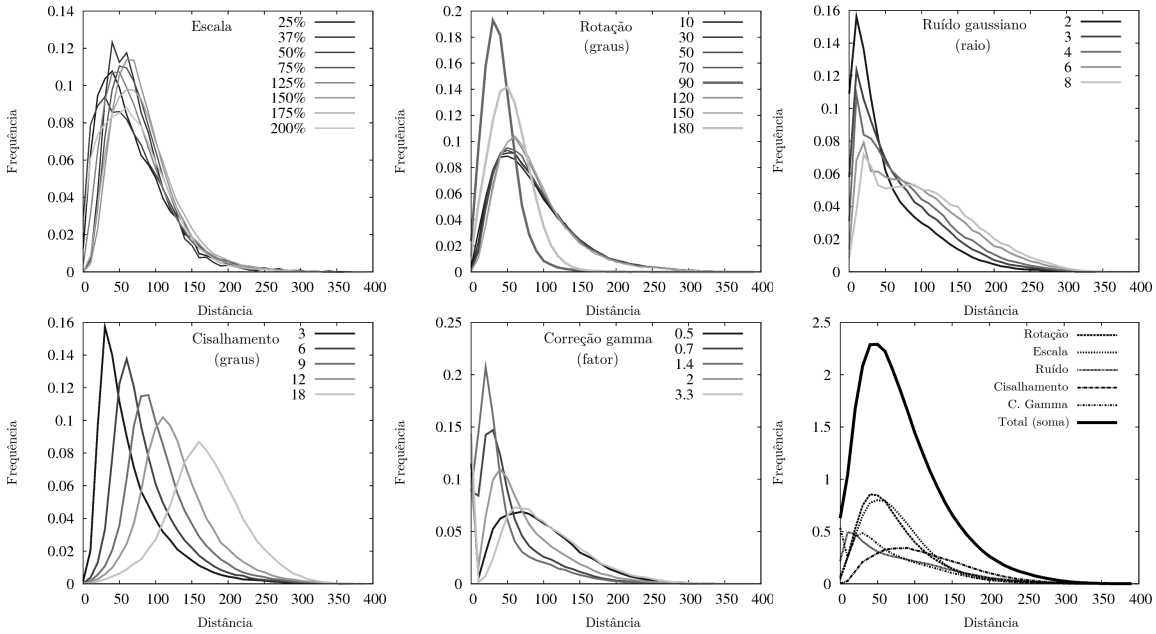

Figura 3.9: Histogramas de distâncias entre vetores de pontos repetidos para várias transformações individuais.

Todos os experimentos usaram a implementação original do SURF, com 64 dimensões[4].

Os mesmos experimentos referentes à análise de escalibilidade feitos para o SIFT também foram feitos para o SURF. O detector Fast Hessian não apresentou problemas de pontos repetidos ambíguos.

A primeira distinção do SIFT para o SURF, obtida experimentalmente, é a diferença de acurácia obtida simplesmente alterando a função de distância Euclidiana (L2) por de quarteirão (L1). Com distância L2, sem base de confusão, e correspondências com 5-vizinhos mais próximos, o SURF obteve um matching score de 70%. Esse valor pulou para 81,8% usando distância L1, ou seja, aumentando mais de 10%. A partir de então, todos os experimentos feitos com o SURF passaram a usar distância de quarteirão.

Apesar de trabalhos apontarem uma tendência, em alta dimensionalidade, de melhor eficácia da distância L1 em relação a L2 (como em [1]), a implementação original do SURF usa distância L2 e a continua usando até sua versão atual (1.0.9). Assim, a melhor eficácia do descritor usando distância L1 é um conhecimento novo.

O SURF foi testado para diferentes tamanhos de base de confusão. O matching score encontrado, por tamanho de base, e por parâmetro K, se encontra na Tabela 3.2. Assim como no SIFT, a acurácia do SURF cai muito lentamente com o aumento da base, permanecendo quase estável na faixa dos 78% para correspondências com 5-vizinhos mais próximos, um pouco inferior se comparado ao SIFT. Isso significa que o SURF também é escalável quanto ao tamanho da base e adequado para ser usado em aplicações de recuperação de semi-réplicas por conteúdo em grandes bases de imagens.

Histogramas de distâncias entre vetores de características foram criados, nos mesmos moldes que no SIFT, tanto para distância L1 como para distância L2. Foram usados 60 bins de tamanho 40 para distância L1 e 40 bins de tamanho 10 para distância L2. Os parâmetros dos histogramas se diferem do SIFT pois a abrangência de valores de distância é diferente. No caso L2, como o SURF possui metade das dimensões do SIFT, a abrangência valores de distâncias é, portanto, reduzida. Porém, no caso L1, a abrangência aumenta, o que poderia influenciar na maior distinguibilidade do SURF para quando a distância L1 for usada.

Percebe-se na Figura 3.10 que a "cauda"do histograma de distâncias entre vetores de pontos repetidos do SURF é maior que do SIFT, aumentado a intersecção com o histograma de distâncias entre vetores de pontos não repetidos, sendo por conseguinte uma causa para menor acurácia do SURF em relação ao SIFT.

Comparando-se as Figuras 3.10a (com distância L2) e 3.10b (com distância L1), nota-se que os histogramas possuem forma semelhante, não sendo evidente uma explicação para a melhor eficácia da distância L1 em relação à L2.

Criou-se, então, histogramas que correlacionam distâncias entre vetores de pontos

[4]http://www.vision.ee.ethz.ch/~surf/download_ac.html (último acesso em 07/07/2011).

Tabela 3.2: Acurácia do descritor SURF para diferentes tamanhos da base de imagens.

Número de vetores de características	700 mil	23 milhões	45 milhões
Número de imagens	225	55 mil	110 mil
Matching score ($K = 5$)	81,8%	78,4%	77,7%
Matching score ($K = 1$)	77,1%	74,0%	73,3%

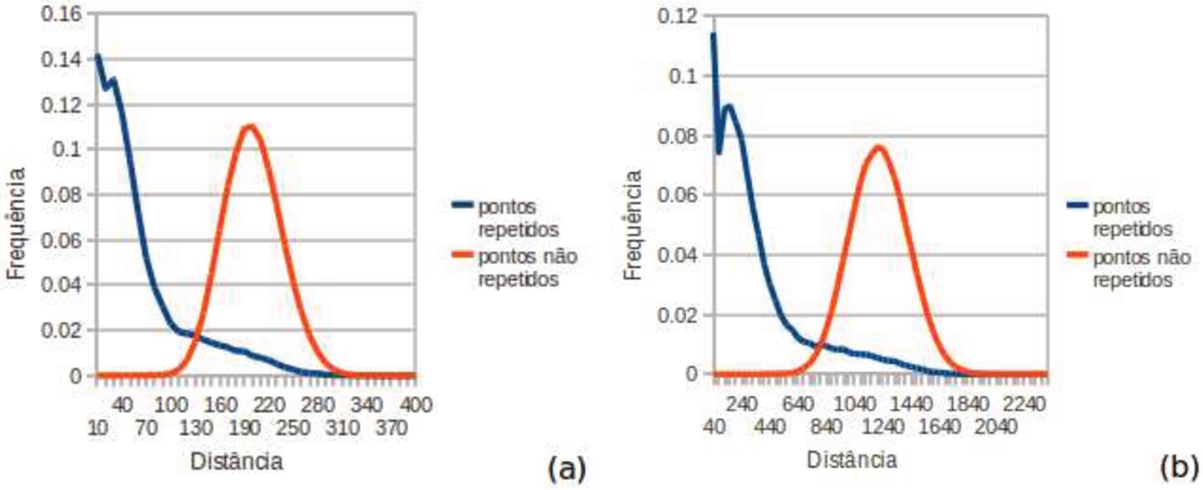

Figura 3.10: Histogramas de distâncias entre vetores de características de pontos repetidos e pontos não repetidos, para o SURF, usando distância Euclidiana (a) e de quarteirão (b).

repetidos e não repetidos, da mesma forma que foi feito para o SIFT. Foi criado um histograma com distância L1 e outro com distância L2, mantendo quantidade e tamanho dos bins.

A Figura 3.11a mostra o histograma de distâncias que correlaciona distâncias entre vetores de pontos repetidos e não repetidos usando distância L2. A Figura 3.11b representa o mesmo histograma em escala logarítmica, que facilita a visualização de valores baixos. Analogamente, a Figura 3.11c mostra o histograma de distâncias que correlaciona distâncias entre vetores de pontos repetidos e não repetidos usando distância L1. A Figura 3.11d representa o mesmo histograma em escala logarítmica. Assim como no SIFT, quando as distâncias entre vetores de pontos repetidos são maiores, as distâncias entre vetores de pontos não repetidos tendem a ser maiores também, contribuindo para que distâncias entre vetores de pontos repetidos permaneçam mínimas e a acurácia do descritor permaneça alta.

As linhas diagonais que partem da origem dos histogramas na Figura 3.11, os dividem em duas partes: a de cima representa os casos em que as distâncias entre vetores de pontos repetidos são menores que as distâncias entre vetores de pontos não repetidos; a de baixo representa os casos contrários. No caso da distância L2, para 71% do histograma, ou mais especificamente, para toda a parte em que a distância entre vetores de pontos repetidos não passa de 70, menos de 5 casos se encontram na parte de baixo do histograma. No caso da distância L1, para 77% do histograma, ou mais especificamente, para toda a parte em que a distância entre vetores de pontos repetidos não passa de 480, menos de 5 casos se encontram na parte de baixo do histograma. Esses resultados são compatíveis com os experimentos de escalabilidade (Tabela 3.2, para distância L1), em que o matching score considerando 5-vizinhos mais próximos são próximos desses valores.

Os histogramas da Figura 3.11 também tornam mais evidente a diferença entre as distâncias Euclidiana (L2) e de quarteirão (L1) entre vetores de características do SURF, evidenciando a distância L1 como a mais apropriada.

3.4 Qual o melhor descritor: SIFT ou SURF?

Considerando somente para as Tabelas 3.1 e 3.2, a resposta para essa pergunta seria claramente SIFT, pois o matching score dele foi maior nos experimentos realizados. Porém, ambos os descritores podem ser usados de forma eficaz em grandes bases de imagens. O SURF, entretanto, posui uma série de vantagens sobre o SIFT que vale a pena ser exposta e que ficou camuflada até agora. Essas vantangens podem até reverter a escolha do SIFT para o SURF.

- Um vetor de características do SURF possui metade das dimensões do SIFT (64

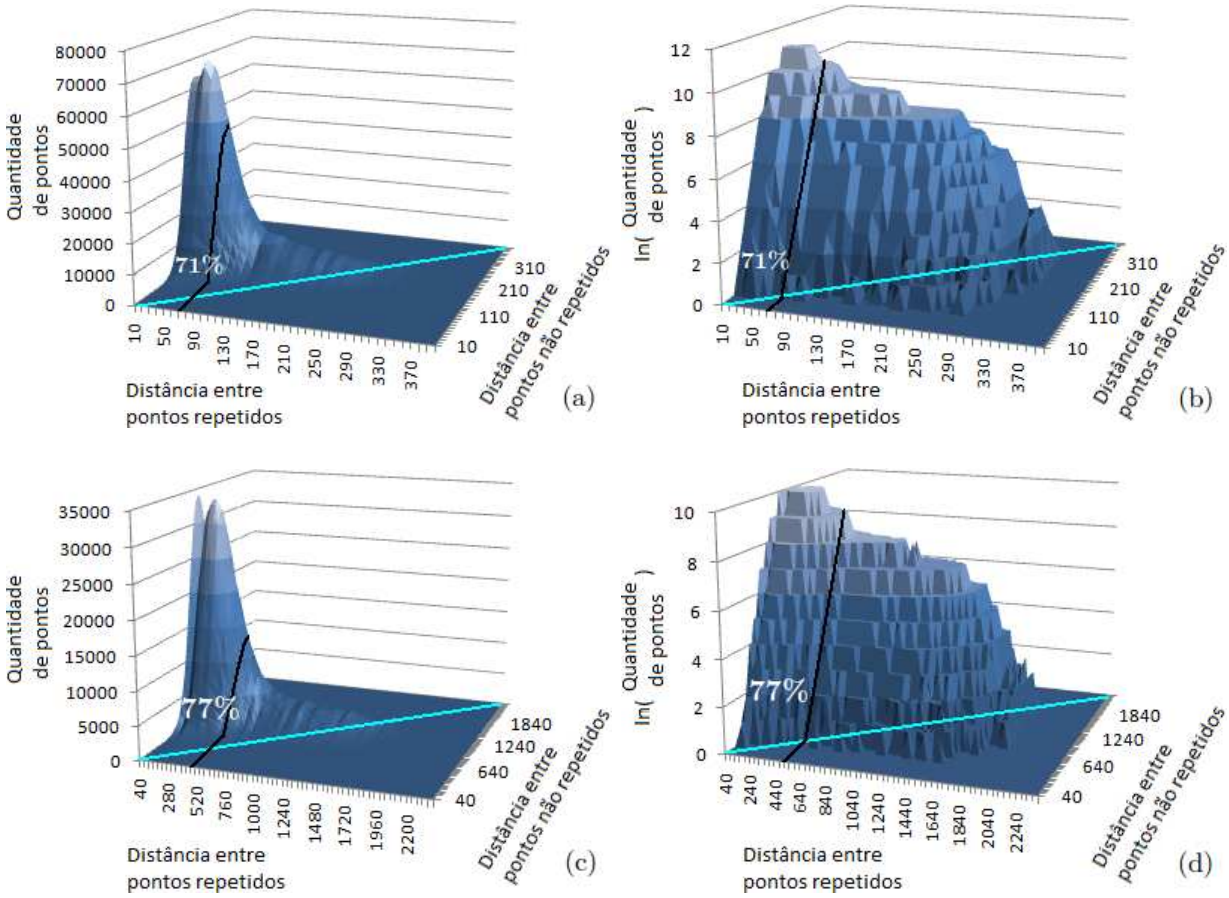

Figura 3.11: Histogramas de distâncias do SURF correlacionando pontos repetidos e não repetidos. Em números absolutos, com distância L2 (a) e L1 (c), e em escala logarítmica, com distância L2 (b) e L1 (d).

contra 128). Sendo assim, vetores do SURF demoram duas vezes menos para terem suas distâncias calculadas, e ocupam duas vezes menos espaço, aumentando a eficiência de tempo de busca e de armazenamento em relação ao SIFT.

- A própria detecção de pontos de interesse e descrição dos mesmos é na ordem de 10 vezes mais rápida para o SURF (como o próprio nome dele diz: Speeded Up Robust Features) do que para o SIFT.

- Este trabalho fez uma análise de robustez entre vetores de pontos repetidos dos descritores SIFT e SURF em alta escala, sem considerar a repetibilidade dos respectivos detectores. Entretanto, na prática, como em uma aplicação de recuperação de semi-réplicas, nada garante que apenas pontos repetidos são usados na busca. Constatou-se que a repetibilidade do DoG (detector do SIFT) cai de 75,9% para 62,1% na remoção de pontos ambíguos, enquanto a repetibilidade do Fast Hessian (detector do SURF) fica em 75,8%. Isso significa que o SURF pode alcançar, na prática, acurácias mais próximas ao SIFT.

Capítulo 4

Detecção de semi-réplicas por Decisão Bayesiana

Como visto no Capítulo 2, técnicas de recuperação de imagens semi-réplicas por conteúdo tendem a usar descritores locais por causa da invariância a transformações e distinguibilidade deles. De alguma forma, uma função de distância é empregada como medida comparativa, seja pela construção de dicionários visuais (eficientes, mas não tão eficazes), ou por algoritmos de votação (eficazes, mas menos eficientes). Apesar de distâncias serem calculadas, o valor delas em si é descartado a cada correspondência encontrada.

Aproveitando o conhecimento adquirido com os experimentos do capítulo anterior, aqui é proposto um novo método de recuperação de imagens semi-réplicas, que faz uso do conhecimento do valor das distâncias entre correspondências de vetores. O método é baseado na teoria de decisão Bayesiana [12] e é compatível com estruturas de indexação para busca aproximada de sistemas de votação. Ele permite recuperar, de forma eficaz e eficiente, uma imagem semi-réplica considerando poucos vetores de características de consulta. Experimentos demonstram que com com cerca de dez de vetores de consulta, chega-se a uma acurácia de 99% na recuperação de semi-réplicas, o que normalmente seria feito com centenas a milhares de vetores no caso de um sistema de votação. Em outras palavras, dada uma imagem de consulta, a velocidade com que uma semi-réplica é encontrada numa base é reduzida na ordem de 10 a 100 vezes, se comparado a sistemas de votação, sem redução de eficácia.

A regra de Bayes já foi aproveitada na literatura na área de CBIR. Em [8] são usados dicionários visuais e decisão Bayesiana para classificação semântica de cenas. Já em [5] Bayes é usado com descriores globais para CBIR com realimentação de relevância. Finalmente, em [26], Bayes é usado para aprimorar a construção de dicionários visuais, investigando a correlação de palavras visuais em imagens. Entretanto, todos esses trabalhos citados usam Bayes de forma e com propósitos bem diferentes do método aqui

desenvolvido.

O método proposto consiste em uma fase de treinamento, que modela probabilisticamente as distâncias entre correspondências de vetores de características, e outra de busca, que aproveita a teoria de decisão Bayesiana para encontrar a imagem original, semi-réplica de uma imagem de consulta.

O descritor usado nos experimentos referentes a este capítulo foi o SIFT, com seu detector padrão (DoG) [21].

4.1 Etapa de Treinamento

O treinamento do método (Algoritmo 1) é o primeiro e importante passo para que a distribuição de probabilidade de distâncias entre correspondências de vetores seja conhecida e que regras de decisão possam ser formuladas a partir do teorema de Bayes [12].

O treinamento consiste, primeiramente, em encontrar correspondências entre vetores de características de imagens de consulta e vetores de uma base constituída por imagens originais e de confusão. As correspondências são feitas por busca de K-vizinhos mais próximos usando, a princípio, $K = 1$ e qualquer função de distância.

Depois de computadas as correspondências, elas são separadas em duas populações: as corretas e as incorretas. Novamente deve-se selecionar um critério de corretude para as correspondências. Dessa vez, ao contrário do Capítulo 3, uma correspondência é correta se, e somente se, o vizinho mais próximo de um vetor de consulta pertence à imagem original da consulta. O critério de corretude de uma correspondência passa a ser, portanto, o mesmo adotado em [18], pois o objetivo é simplesmente recuperar a imagem correta, sem se preocupar se o ponto é repetido ou não.

Calcula-se, então, para cada uma dessas duas populações, um histograma de distâncias entre os vetores de consulta e suas correspondências. Consideraram-se histogramas com 50 bins de tamanho 7. Comparando-se ao Capítulo 3, nota-se a diminuição do tamanho dos bins dos histogramas de 15 para 7. Essa diminuição deve-se ao fato que as distâncias entre correspondências são menores que as distâncias entre vetores de pontos repetidos e ainda menores que as distâncias entre vetores quaisquer.

Em seguida, cada histograma é ajustado por uma função de densidade de probabilidade (fdp), usada posteriormente na fase de busca do método. Essa função também serve para suavisar o histograma, eliminando ruídos.

A frequência de correspondências corretas no total de correspondências (probabilidade a priori de correspondência correta) deve ser armazenada para etapa de busca Bayesiana.

Algoritmo 1 – Etapa de Treinamento

Considere os vetores de características de todas imagens previamente computados.

Entrada: Ponteiro para bases com imagens originais $BOriginal$, de confusão $BConfusao$ e de consulta $BConsulta$, função de distância D, função de ajuste A.

Saída: Parâmetros de função ajustada para histograma de correspondências corretas $ParCorreto$ e incorretas $ParIncorreto$, probabilidade a priori de correspondência correta $ProbCorreto$.

Auxiliares: Vetor de características v, distância $dist$, histogramas de correspondências corretas $HistCorreto$ e incorretas $HistIncorreto$, índice i, contadores $ContCorreto$ e $ContIncorreto$.

1. *Para todo* v *em* $BConsulta$
2. $\quad \{dist, i\} \leftarrow vizinho\_mais\_proximo(v, BOriginal \cup BConfusao, D)$
3. $\quad$ *Se* $i \in BOriginal$
4. $\quad\quad incrementa\_histograma(HistCorreto, dist)$
5. $\quad\quad ContCorreto \leftarrow ContCorreto + 1$
6. $\quad$ *Senão*
7. $\quad\quad incrementa\_histograma(HistIncorreto, dist)$
8. $\quad\quad ContIncorreto \leftarrow ContIncorreto + 1$
9. $ProbCorreto \leftarrow \frac{ContCorreto}{ContCorreto+ContIncorreto}$
10. $ParCorreto \leftarrow ajusta\_em\_fdp(HistCorreto, A)$
11. $ParIncorreto \leftarrow ajusta\_em\_fdp(HistInorreto, A)$

4.1.1 Correspondências exatas

Como visto no capítulo anterior, uma função de distância adequada para ser usada com o SIFT é a Euclidiana (L2). Calculando-se a distância Euclidiana entre um vetor de característica de consulta e todos os demais vetores da base, encontra-se exatamente o vizinho mais próximo do vetor de consulta, formando uma correspondência exata na métrica L2.

A Figura 4.1 mostra os histogramas de correspondências corretas (curvas pretas das Figuras 4.1a e 4.1b) e de correspondências incorretas (curvas pretas das Figuras 4.1c e 4.1d), criados a partir de busca por vizinho mais próximo com distância Euclidiana. Nota-se que o histograma de correspondências corretas têm uma curva muito mais próxima de zero do que aquele de correspondências incorretas, o que significa que a distância entre os vetores de características para o primeiro caso têm valores frequentemente inferiores aos valores para o segundo caso, como se é esperado.

As curvas da Figura 4.1 consideram 110.000 imagens de confusão, 78 imagens originais

e uma imagem de consulta por imagem original (metade daquelas usadas nos experimentos relativos à escalabilidade de descritores no Capítulo 3, porém sem incluir transformações que contenham somente recorte).

4.1.2 Correspondências aproximadas

Como visto no Capítulo 2, do ponto de vista computacional, o cálculo da distância entre vetores de características de consulta e todos os demais vetores da base é muito caro e pode levar várias horas, dependendo do tamanho da base. Existem técnicas de busca por vizinho mais próximo bem mais eficientes, porém menos eficazes, que fazem um número bem menor de operações na base, como em [11,38,40]. Qualquer técnica baseada em votação pode ser adaptada para o método proposto, como por exemplo, a técnica baseada em Multicurves [40], que foi usada nos experimentos devido ao alto desempenho dela em [40][1]. As correspondências encontradas por essas técnicas são aproximadas, pois nem todos os vetores da base são checados em uma consulta.

A Figura 4.2 mostra os histogramas de correspondências corretas (curvas pretas das Figuras 4.2a e 4.2b) e de correspondências incorretas (curvas pretas das Figuras 4.2c e 4.2d), criados a partir de busca por vizinho mais próximo com Multicurves. Note que os histogramas variam pouco em relação àqueles da Figura 4.1, podendo todos serem tratados de forma semelhante.

As curvas das Figuras 4.2 e 4.1 consideram as mesmas imagens.

4.1.3 Modelo estatístico

Independentemente da função de distânica usada, os histogramas computados podem ser modelados por uma distribuição de probabilidade. No caso do SIFT, como pode ser visto nas Figuras 4.1 e 4.2 (curvas pretas), os histogramas de distância são assimétricos. Uma distribuição simétrica frequentemente usada, como a normal, não se ajusta bem aos dados (curvas cinzas das Figuras 4.1b, 4.1d, 4.2b, 4.2d). Uma distribuição não simétrica seria melhor. Várias distribuições foram testadas, dentre elas: Chi, Chi-quadrado, Weibull e log-normal. A Chi acabou selecionada (curvas cinzas das Figuras 4.1a, 4.1c, 4.2a, 4.2c). Não só a distribuição Chi teve o melhor ajuste global, com parâmetros mais satisfatórios, mas também tem a explicação mais satisfatória quando a distância L2 é usada. Uma vez que a distribuição Chi é dada pela raiz quadrada da soma dos quadrados de normais independentes, ela pode ser explicada se for considerado que cada dimensão dos vetores de características do SIFT tem uma distribuição normal independente.

[1] Agradecemos a Fernando Akune por fornecer os índices Multicurves dos vetores de características da base, bem como o programa para que a busca fosse efetuada.

O ajuste dos histogramas foi calculado pelo método de mínimos quadrados não-linear [16]. Dado um parâmetro inicial, esse método minimiza, iterativamente, a diferença quadrática entre a função de ajuste e o dado real (os valores dos histogramas).

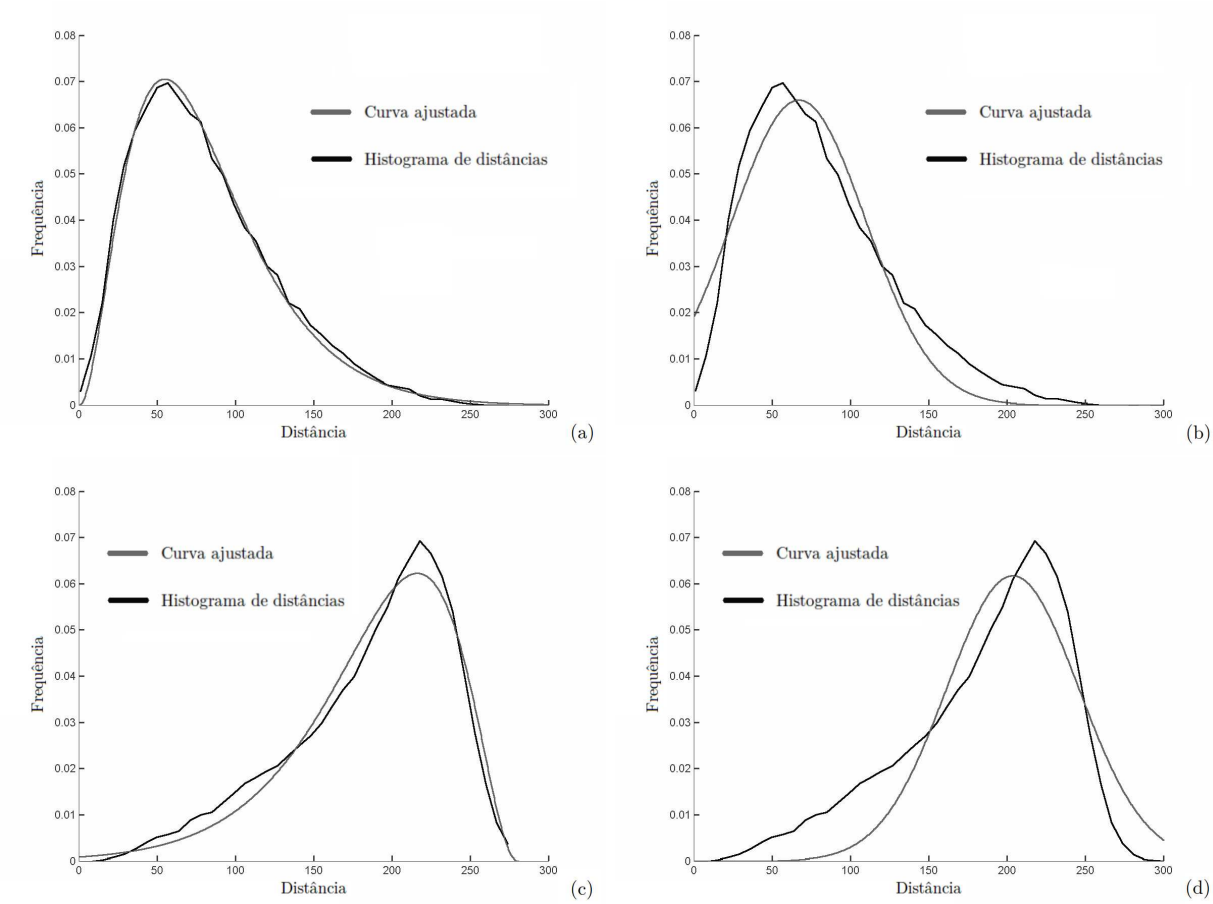

Figura 4.1: Os histogramas de distâncias corretas (a e b) e incorretas (c e d), ajustados por uma distribuição Chi (a e c) e uma Normal (c e d). A partir de correspondências exatas com distância Euclidiana.

4.2 Busca Bayesiana

Combinando a teoria da decisão Bayesiana [12] e o modelo estatístico para distâncias entre correspondências, pode-se recuperar uma imagem semi-réplica com precisão elevada usando-se poucos vetores de características da imagem de consulta. O desenvolvimento matemático a seguir vale tanto para o modelo com busca exata como para busca aproximada.

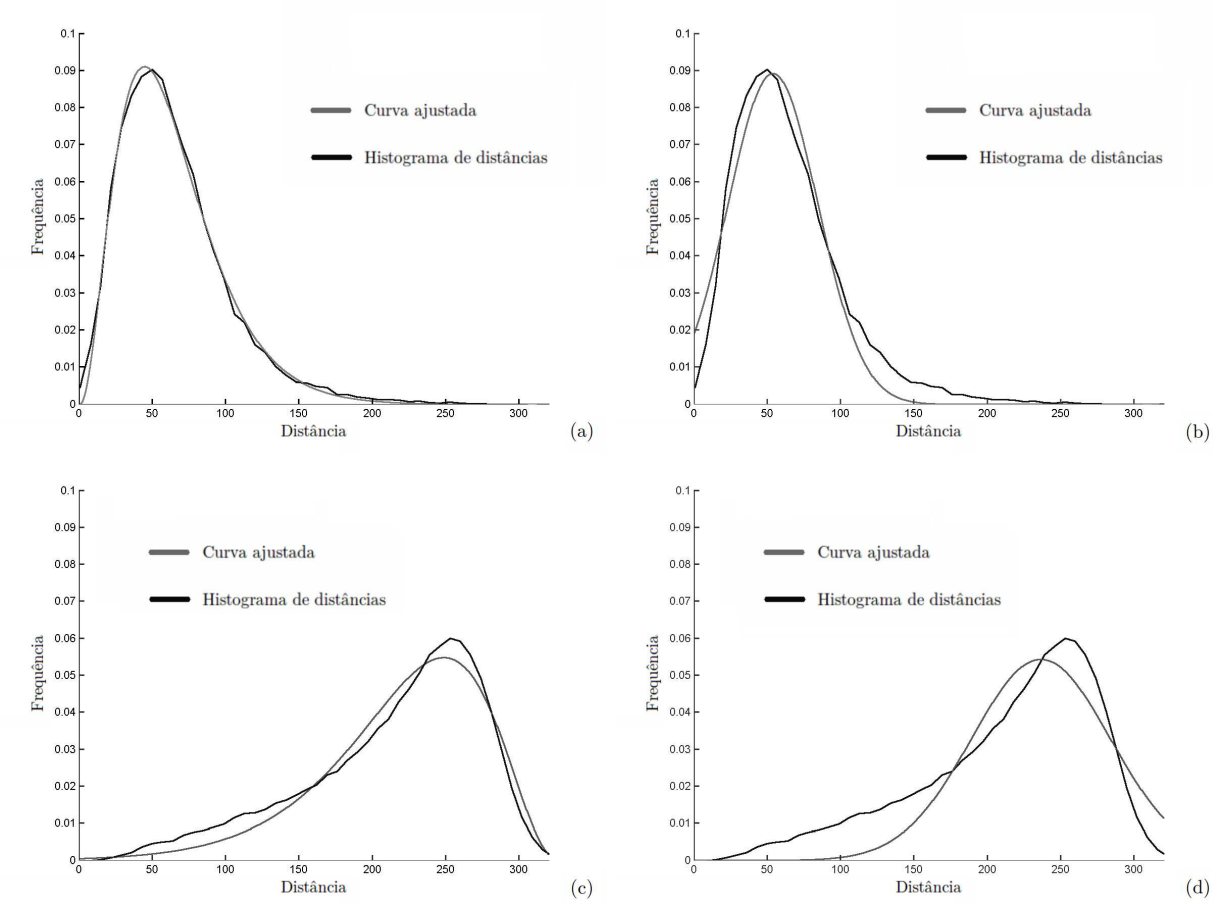

Figura 4.2: Os histogramas de distâncias corretas (a e b) e incorretas (c e d), ajustados por uma distribuição Chi (a e c) e uma Normal (c e d). A partir de correspondências aproximadas ancontradas pelo método Multicurves.

Primeiramente, as funções de densidade de probabilidade encontradas com modelo estatístico são discretizadas para o domínio dos números inteiros. Esse passo tanto facilita a implementação (os valores de probabilidade podem ser pré-calculados e armazenados em memória) como facilita o formalismo matemático, lidando com probabilidades discretas e não contínuas.

Considere $P(X)$ a probabilidade a priori de que uma correspondência esteja correta (isto é, a probabilidade de que o vizinho mais próximo de um vetor de características de consulta irá se corresponder a um vetor de sua imagem original), que pode ser obtido durante a fase de treinamento.

Sendo assim, $P(\overline{X})$ é a probabilidade a priori de que uma correspondência é incorreta e $P(X) = 1 - P(\overline{X})$.

$P(D|X)$ e $P(D|\overline{X})$ são as probabilidades condicionais de se encontrar uma distância D, dado que a correspondência é correta ou incorreta, respectivamente. Utilizando o modelo estatístico obtido para ajustar os histogramas de distâncias, essas probabilidades podem ser calculadas analiticamente.

A taxa de verossimilhança de que uma única correspondência i seja correta é dada pela Equação 4.1.

$$L_i = \frac{P(D_i|X) + \epsilon}{P(D_i|\overline{X}) + \epsilon} \tag{4.1}$$

ϵ é um pequeno número para evitar divisão por zero.

A probabilidade a posteriori de que uma imagem j da base seja correta após n correspondências feitas entre ela e amostras aleatórias de vetores da imagem de consulta é dada pela Equação 4.2. A aleatoriedade das amostras é importante para evitar correspondências com distâncias dependentes entre si, o que tornaria a Equação 4.2 inválida. Por exemplo, sabe-se que correspondências entre vetores de características de pontos de interesse em áreas de borda de um recorte possuem distâncias maiores, como visto no Capítulo 3. Se somente vetores de pontos em áreas de borda fossem selecionados, a probabilidade a posteriori de que um imagem semi-réplica da base fosse correta seria bem inferior ao esperado.

$$P_j(X|D_1 \cap D_2 \ldots \cap D_n) = \frac{\prod_{i=1}^{n} L_i \times P(X)}{\prod_{i=1}^{n} L_i \times P(X) + P(\overline{X})} \tag{4.2}$$

Pode-se, então, definir um limite superior $\mathcal{T}$ para $P_j(X|D_1 \cap D_2 \ldots \cap D_n)$ para o qual não mais amostras são tomadas da consulta, de maneira que a imagem j seja recuperada com precisão desejável e com poucas buscas por vizinho mais próximo na base. Em outras palavras, se $\exists j | P_j(X|D_1 \cap D_2 \ldots \cap D_n) > \mathcal{T}$ então a imagem j é considerada correta e a busca para.

4.2.1 Detecção de consultas inválidas

Em algumas aplicações a imagem de consulta pode não ter uma correspondente semi-réplica na base de dados (consultas inválidas). Assim, a busca Bayesiana proposta até agora não pode detectar uma consulta inválida e vários, ou até mesmo todos os vetores de características de uma imagem de consulta, seriam processados para busca retornar uma imagem incorreta.

Para resolver esse problema, considere $P(\overline{Q})$ a probabilidade a priori de que a consulta seja inválida. $P(\overline{Q})$ é dependente de aplicação, mas seu valor não necessita ser extremamente fiel à realidade, pois o resultado da detecção de consultas inválidas é mais influenciado pela taxa de verossimilhança (Equação 4.4) do que por ele.

Então $P(Q)$ é a probabilidade a priori que a consulta tem uma correspondente semi-réplica na base de dados.

$P(D|\overline{Q})$ e $P(D|Q)$ são as probabilidades condicionais de se encontrar uma distância D, dado que a consulta é inválida ou válida, respectivamente.

$P(D|\overline{Q})$ pode ser obtido durante a fase de treinamento e possui exatamente o mesmo valor que $P(D|\overline{X})$, já que todas correspondências encontradas a partir de uma imagem de consulta inválida são falsas.

Para que a igualdade $P(D|\overline{Q}) = P(D|\overline{X})$ seja verdadeira, é importante que a fase de treinamento do método permaneça sem mudanças, ou seja, que nenhuma consulta inválida seja inserida nela. Sendo assim, pode-se afirmar também que o valor de $P(D|Q)$ é igual ao valor anterior de $P(D)$, quando consultas inválidas não eram consideradas no modelo. Portanto, expandindo $P(D|Q)$ chega-se à Equação 4.3.

$$P(D|Q) = P(X) \times P(D|X) + P(\overline{X}) \times P(D|\overline{X}) \tag{4.3}$$

A igualdade da Equação 4.3 é de valores, somente, causada pelo artifício de se manter a fase de treinamento do método sem inserção de consultas inválidas.

A taxa de verossimilhança de que uma única correspondência i indique que uma consulta seja inválida é dada pela Equação 4.4.

$$LQ_i = \frac{P(D_i|\overline{X}) + \epsilon}{P(X) \times P(D_i|X) + P(\overline{X}) \times P(D_i|\overline{X}) + \epsilon} \tag{4.4}$$

ϵ é um pequeno número para evitar divisão por zero.

A probabilidade que a consulta seja inválida depois de N correspondências feitas entre vetores da base de dados e amostras aleatórias de vetores da imagem de consulta é dada pela Equação 4.5.

$$P(\overline{Q}|D_1 \cap D_2 \ldots \cap D_N) = \frac{\prod_{i=1}^{N} LQ_i \times P(\overline{Q})}{\prod_{i=1}^{N} LQ_i \times P(\overline{Q}) + P(Q)} \tag{4.5}$$

Pode-se, então, definir um limite inferior τ para $P(Q|D_1 \cap D_2 \ldots \cap D_N)$ para o qual não mais amostras são tomadas e a imagem de consulta seja rejeitada (assumida ser inválida). Note que $P(Q|D_1 \cap D_2 \ldots \cap D_N)$ é um valor global, enquanto $P_j(X|D_1 \cap D_2 \ldots \cap D_n)$ é calculado para cada imagem j da base.

4.2.2 Tratamento de caso: distância zero

Analisando os histogramas ajustados pela função Chi nas Figuras 4.1 e 4.1, percebe-se que os valores de densidade de probabilidade da Chi são nulos tanto para o caso de correspondências corretas como para o caso de correspondências incorretas (nesse último caso o valor é matematicamente maior que zero, porém não suficientemente grande para ser maior que zero com precisão de ponto flutuante).

Sendo assim, uma distância zero não tem importância estatística e apenas diminuiria a eficiência e eficácia do método proposto. Entretanto, como ela ocorre somente em casos particulares (imagens réplicas, imagens somente com recorte ou, como visto no Capítulo 3, imagens com mudanças de luminosidade por correção gamma) ela também pode ser tratada de forma particular e bem simples: se uma distância zero for encontrada entre dois vetores, pode-se inferir que a região definida pelos pontos de interesses referentes a esses vetores são exatamente as mesmas. Isso significa que as imagens referentes a esses pontos são semi-réplicas. Pode-se então adicionar a seguinte exceção ao método: *Se a distância entre um vetor de características de uma imagem de consulta e um vetor consultado da base for zero, então a busca para e a imagem referente ao vetor consultado é retornada.*

Finalmente o algoritmo final para busca Bayesiana é descrito (Algoritmo 2).

4.3 Resultados

Aqui são mostrados os resultados de um conjunto de três experimentos para validar o método apresentado de recuperação de semi-réplicas baseado na teoria de decisão Bayesiana:

1. Avaliação do método usando correspondências exatas, sem detecção de consultas inválidas (todas consultas são verdadeiras).

2. Avaliação do método usando correspondências aproximadas por indexação de Multicurves [40], também sem detecção de consultas inválidas.

3. Avaliação do método em relação a detecção de consultas inválidas e tratamento de distância zero.

Algoritmo 2 – BUSCA BAYESIANA

Considere os vetores de características da base de imagens previamente computados, bem como as probabilidades condicionais de correspondência correta e incorreta (a partir dos parâmetros encontrados no Algotitmo 1).

ENTRADA: Imagem de Consulta Q, ponteiro para base de imagens $Base$, probabilidades condicionais de correspondência correta $P(D|X)$ e incorreta $P(D|\overline{X})$, probabilidades a priori de correspondência correta $P(X)$ e de consulta inválida $P(\overline{Q})$, limiares de aceitação de imagem $\mathcal{T}$ e rejeição de consulta τ, função de distância D.

SAÍDA: Imagem semi-réplica ND.

AUXILIARES: Conjunto de vetores de características V, vetor v, distância $dist$, índice i, imagem I, taxas de verossimilhança das imagens L e da consulta LQ, probabilidade de distancia $P(D|Q)$, probabilidades a posteriori de imagem correta $P(X|L)$ e de consulta inválida $P(\overline{Q}|LQ)$.

1. $ND \leftarrow \emptyset$
2. $V \leftarrow extracao\_de\_caracteristicas(Q)$
3. *Enquanto* $v \leftarrow remove\_vetor\_aleatorio(V)$
4. $\quad \{dist, i\} \leftarrow vizinho\_mais\_proximo(V, Base, D)$
5. $\quad I \leftarrow$ *imagem com vetor* i
6. $\quad$ *Se* $dist = 0$
7. $\quad\quad ND \leftarrow I$
8. $\quad\quad$ *Para*
9. $\quad$ *Se* $L[I] = \emptyset$
10. $\quad\quad L[I] \leftarrow 1$
11. $\quad\quad P(X|L)[I] \leftarrow P(X)$
12. $\quad L[I] \leftarrow L[I] \times \frac{P(D|X)[dist]}{P(D|\overline{X})[dist]}$
13. $\quad P(X|L)[I] \leftarrow \frac{L[I] \times P(X)}{L[I] \times P(X) + 1 - P(X)}$
14. $\quad P(D|Q) \leftarrow P(D|X)[dist] \times P(X) + P(D|\overline{X})[dist] \times (1 - P(X))$
15. $\quad LQ \leftarrow LQ \times \frac{P(D|\overline{X})[dist]}{P(D)}$
16. $\quad P(\overline{Q}|LQ) \leftarrow \frac{LQ \times P(\overline{Q})}{LQ \times P(\overline{Q}) + 1 - P(\overline{Q})}$
17. $\quad$ *Se* $P(X|L)[I] > \mathcal{T}$
18. $\quad\quad ND \leftarrow I$
19. $\quad\quad$ *Para*
20. $\quad$ *Se* $P(\overline{Q}|LQ) > \tau$
21. $\quad\quad$ *Para*
22.

Os dois primeiros experimentos servem para avaliar a eficiência do método proposto, mostrando que poucos vetores de características de consulta são necessários para se encontrar uma imagem semi-réplica, mantendo alta eficácia (até $99,3\%$). Eles também comparam os modelos chi e normal, mostrando que o modelo chi é mais fidedigno à distribuição de distâncias entre vetores do SIFT.

O segundo experimento também mostra que o uso de técnicas de indexação, com correspondências aproximadas, em particular Multicurves, é totalmente compatível com o modelo proposto, mantendo alta eficácia (até $99,1\%$). O que significa que essas técnicas, que aumentam a eficiência de sistemas de recuperação de semi-réplicas por votação, podem ser ainda mais eficientes se aliadas ao método baseado na teoria de decisão Bayesiana.

O terceiro experimento tem a finalidade de mostrar dois casos especiais: que o método proposto rejeita, de forma eficaz e eficiente, imagens de consulta que não estão na base, e que pares de vetores com distância zero retornam corretamente imagens réplicas ou semi-réplicas com apenas recorte.

Como o método proposto não é determinístico, todos experimentos foram executados 10 vezes e a média de cada resultado foi tomada. Por questões de praticidade, as distâncias entre as correspondências foram todas pré-calculadas e armazenadas.

4.3.1 Busca Bayesiana com correspondências exatas

Testou-se a busca por decisão Bayesiana com correspondências exatas, descrita anteriormente, com 110.000 imagens da base de confusão. As imagens de consulta foram geradas por transformações de 156 imagens originais, as mesmas dos experimentos relativos à escalabilidade de descritores no Capítulo 3, porém sem incluir transformações que contenham somente recorte.

A metade das imagens de consulta foi usada para o teste (busca), enquanto a outra metade foi usada para o treinamento. Foi tomado cuidado para que tanto a parte usada para consulta como a parte usada para treinamento representassem, em proporção semelhante, todos os tipos de transformações da base de consulta. Além disso, como visto no Capítulo 3, o treinamento não deve sofrer grandes alterações se a base fosse trocada.

A distância Euclidiana foi usada para encontar o vizinho mais próximo dos vetores de consulta. As correspondências formadas foram exatas, pois, para cada vetor de consulta, foram calculadas distâncias entre ele e todos vetores da base.

A probabilidade a priori ($P(X)$) encontrada no treinamento foi de 45,0%.

Como se pode ver na Tabela 4.1, a busca Bayesiana apresentou alta precisão utilizando, em média, por volta de 10 (entre centenas ou milhares) de vetores de características por consulta. Em outras palavras, o método proposto reduz na ordem de 10 a 100 vezes o tempo de processamento de uma busca por semi-réplicas por votação. Ainda assim a

Tabela 4.1: Acurácia e número médio de amostras necessárias para busca Bayesiana usando distância Euclidiana.

	Limiar $\mathcal{T}$	90,00%	99,00%	99,90%
Ajuste Normal	Média de amostras	2,5	3,9	6,2
Ajuste Normal	Acurácia	85,5%	90,8%	97,3%
Ajuste Chi	Média de amostras	3,8	6,8	10,2
Ajuste Chi	Acurácia	93,4%	98,7%	99,3%

acurácia chega a superar 99%.

Mesmo que com a distribuição normal menos amostras aleatórias de vetores de características foram usadas para um mesmo limiar, sua precisão é notavelmente pior do que com a distribuição Chi. O menor número de amostras usadas com a distribuição normal é devido à baixa intersecção entre as curvas que ajustam os histogramas de distância de correspondências corretas e incorretas. Porém, usando a distribuição normal, as caudas dos histogramas não ficam bem representadas e uma quantidade de correspondências incorretas passa como correta, o que diminui sua acurácia. Esse problema não ocorre quando a curva Chi é usada, possuindo uma acurácia próxima do esperado, ou seja, próxima do limiar de probabilidade para o qual a imagem recuperada seja correta.

4.3.2 Busca Bayesiana com correspondências aproximadas

Testou-se a busca por decisão Bayesiana com correspondências exatas, descrita anteriormente, com as mesmas imagens originais, de confusão e de consulta.

A metade das imagens de consulta foi usada para o teste (busca), enquanto a outra metade foi usada para o treinamento. A técnica baseada em Multicurves (detalhada no Capítulo 2) foi usada para encontrar o vizinho mais próximo de um vetor de consulta. Foram usados 8 sub-índices com 512 elementos examinados em cada sub-índice.

A probabilidade a priori $P(X)$ encontrada no treinamento foi de 27,2%, notavelmente menor que a probabilidade a priori obtidada pelo método com correspondências exatas.

Como se pode ver na Tabela 4.2, a busca Bayesiana com correspondências aproximadas apresentou precisão comparável à busca com correspondências exatas, utilizando, em média, um pouco mais de amostras vetores de consulta. Essa quantidadade a mais de amostras, entretanto, aumenta substancialmente menos o tempo de processamento do método se comparada com a diminuição de tempo causada pelo emprego da busca aproximada. Isso significa que a busca Bayesiana com correspondências aproximadas aprimora, em termos de eficiência, a busca Bayesiana com correspondências exatas, assim como sistemas baseados em votação são aprimorados, em termos de eficiência, com correspondências aproximadas. E, não menos importante, a busca Bayesiana com correspondências aproxi-

Tabela 4.2: Acurácia e número médio de amostras necessárias para busca Bayesiana usando Multicurves.

	Limiar $\mathcal{T}$	90,00%	99,00%	99,90%
Ajuste Normal	Média de amostras	2,9	4,4	6,5
Ajuste Normal	Acurácia	86,1%	91,4%	98,0%
Ajuste Chi	Média de amostras	4,9	8,8	11,1
Ajuste Chi	Acurácia	92,9%	98,0%	99,1%

madas, por usar poucos vetores de características de consulta, aprimora a busca baseada em votação com correspondências aproximadas.

Medição de tempo Para deixar mais clara a eficiência do método, a Tabela 4.3 apresenta uma estimativa de tempo de consulta para detecção de semi-réplicas.

Como os experimentos usaram, por questões de praticidade, distâncias pré-calculadas na etapa de busca, os números da Tabela 4.3 não apresentam dados reais de consulta, mas os representam bem. Eles foram obtidos considerando 1000 vetores de características de consulta, compatível com a quantidade média de vetores em uma imagem. Para cada um desses vetores, foram feitos dois experimentos, com seus tempos de execução medidos. O primeiro calcula a distância L2 entre cada um dos vetores e todos 70 milhões de vetores da base de confusão, como em um sistema de votação com busca exata. O segundo calcula a distância, usando Multicurves, entre cada um dos vetores com os vetores da base de confusão, como em um sistema de votação com busca aproximada. O primeiro experimento foi executado em torno de 2200 segundos, enquanto o segundo em 260 segundos. O tempo de detecção e descrição de pontos de interesse do SIFT demora cerca de 1 segundo para uma imagem. Sendo assim, lembrando que a etapa de consulta, ou fase *online*, de um sistema de detecção de semi-réplicas por votação consiste na detecção e descrição de pontos de interesse, seguido de uma busca por vizinho mais próximo, os valores obtidos representam bem o tempo de consulta em sistemas de votação.

Nota-se a diferença na ordem de grandeza entre os tempos de busca exata e aproximada. Além disso, mais um detalhe na implementação de medição de tempo fortalece a vantagem da busca aproximada: o cálculo de distância L2 entre vetores de características foi feito em 12 threads com paralelismo total, enquanto o cálculo de distância com Multicurves foi feito sem paralelismo. Entretanto, o Multicurves é paralelizável [37], o que diminuiria ainda mais os valores da Tabela 4.3.

Agora podemos estimar o tempo de consulta usando o método proposto de busca Bayesiana aliado a uma técnica de busca aproximada, como o Multicurves. Os cálculos de probabilidade efetuados no método consomem tempo despresível em relação às demais operações. Sendo assim, o tempo de consulta consiste basicamente na soma do tempo de

Tabela 4.3: Estimativas de tempo de busca para uma imagem de consulta com 1000 pontos de interesse (em segundos).

Método	Tempo (s)
Votação com L2	2200
Votação com multicurves	260
Bayesian com multicurves	4

detecção e descrição de pontos de interesse da imagem de consulta e o tempo de busca por vizinho mais próximo para cada vetor de característica usado. Como o tempo total de busca é proporcional à quantidade de amostras de vetores de consulta, para uma imagem com 1000 vetores, se 11 deles forem amostrados (de acordo com a tabela 4.2), em pouco menos de 3 segundos todas as operações com a base de dados são feitas. Somando-se o tempo de detecção e descrição do SIFT , em torno de 4 segundos uma semi-réplica é retornada.

A máquina usada para medição de tempo tem placa mãe Intel S5520SC; processador Intel Xeon X5670, 2.93Ghz, com 6 cores, 12 threads e 12Mb de cache; possui 12 Gb de RAM (6 x 2Gb DDR3 1333); 4 discos 1.5 TB SATA II (onde estão armazenados os dados).

4.3.3 Casos especiais: detecção de consultas inválidas e distância zero

Incluiu-se a detecção de consultas inválidas na busca, usando Multicurves e modelo Chi. Estabeleceram-se a probabilidade a priori de consulta inválida $P(\overline{Q})$ em 50% e o limite inferior de rejeição (τ) em 0,1%.

Para testar consultas com imagens originais na base (consultas válidas), o mesmo conjunto de imagens de consulta dos experimentos anteriores foi usado. Apenas 0,1% das consultas foram erroneamente rejeitadas, como o esperado.

Para testar consultas sem imagens originais na base (consultas inválidas), um conjunto de 44 imagens foram usadas. 1,5% das consultas foram erroneamente aceitas. Na média, 25,7 amostras de vetores de características das imagens de consulta foram tomadas antes da consulta ser rejeitada.

Quando imagens somente com recorte foram inseridas na consulta, 99,9% das vezes foi retornada a imagem original correta, 95% delas em uma iteração. Por motivos triviais, não foram registradas imagens incorretas com alguma distância entre vetores de características de valor zero. Não foram testadas imagens de consulta com alguma réplica na base pois a acurácia seria obviamente 100%.

Com esse experimento, conclui-se que imagens que não pertencem à base podem ser usadas como consulta, pois o método proposto detecta de forma eficaz e eficiente se a

consulta possui ou não semi-réplica na base. Também conclui-se que casos com distância zero podem ser tratados de forma simples e eficaz.

Capítulo 5

Discussão

Este capítulo retoma as contribuições deste trabalho, enfatizando os resultados obtidos, bem como apresenta sugestões para trabalhos futuros.

5.1 Contribuições

Este trabalho contribuiu para o estado da arte na área de descritores locais aplicados à recuperação de imagens semi-réplicas nos seguintes pontos:

1. Estudo da influência do tamanho da base de imagens na qualidade de descritores locais, no contexto de detecção de semi-réplicas: confirmou-se que os descritores locais SIFT e SURF são invariantes e distinguíveis o suficiente para serem usados em aplicações com grande quantidade de dados, mantendo matching score de 73% a 92% para testes feitos com 1 milhão a 70 milhões de pontos de interesse. Esses descritores já são amplamente usados em detecção de semi-réplicas, como visto no Capítulo 2. Sendo assim, o resultado era esperado.

2. Estudo da influência da função de distância na qualidade do descritor: descobriu-se que o SURF é mais preciso quando a distância L1 é usada, em detrimento da distância L2, confirmando a importância da função de distância em aplicações de detecção de semi-réplicas.

3. Análise refinada do comportamento das distâncias entre vetores de características, por meio de histogramas de distância: constatou-se, tanto para o SIFT como para o SURF, que histogramas de distâncias entre vetores de pontos repetidos (corretos no ponto de vista do detector) e não repetidos (incorretos) possuem pequena intersecção. Distâncias entre vetores de pontos repetidos são bem menores, justificando a qualidade dos descritores, quanto à invariância e à distinguibilidade. Também

foi mostrado, com histogramas que correlacionam distâncias entre vetores de pontos repetidos e não repetidos, que, com distâncias maiores entre vetores de pontos repetidos, distâncias entre vetores de pontos não repetidos também são maiores. Essa descoberta contribui para escalabilidade dos descritores. Especificamente para o SIFT, um estudo com diversas transformações foi feito, constatando um padrão de comportamento de histogramas de distâncias, que pode ser modelado estatisticamente por uma função de densidade de probabilidade assimétrica, como a Chi.

4. Proposição de um novo método eficiente e eficaz de detecção de semi-réplicas, baseado na teoria de decisão Bayesiana e na modelagem estatística dos histogramas de distância que diminui a quantidade de consultas à base de dados para recuperar uma imagem semi-réplica. A redução parte de centenas ou milhares de vetores de características por imagem de consulta, comumente presentes em imagens, para em torno de 10, sem redução de eficácia (mantendo-se acima de 99%).

5. Expansão do método proposto para identificar imagens com nenhuma semi-réplica na base de dados: consultas inválidas (sem semi-réplica na base) são detectadas corretamente em 98,5 % das vezes, com 0,1% de falsos negativos.

6. Integração do método proposto com estruturas de indexação para melhorar ainda mais sua eficiência: nos experimentos foi usada a técnica Multicurves para indexação de vetores e busca aproximada por vizinho mais próximo. Entretanto, qualquer outra técnica similar pode ser usada.

5.2 Trabalhos Futuros

As seguintes sugestões para trabalhos futuros são dadas:

- Repetir os experimetos feitos com SIFT e SURF no Capítulo 3 com outros descritores, como GLOH [25] e PCA-SIFT [14], incluindo testes com o sentido de orientação dos pontos de interesse, para evitar pontos repetidos ambíguos.

- Incluir transformações mais pesadas nas imagens, como mudanças de perspectiva, reforçando ou não a tese de que as distribuições de distância entre vetores de características são robustas em relação a base de consulta.

- Construção de um arcabouço baseado no método proposto de detecção de semi-réplicas, possibilitando medição precisa de tempo de execução de consultas (considerando que os experimentos realizados neste trabalho usaram distâncias pré-calculadas entre vetores de características de consulta e vetores da base), bem como o oferecimento do método para comunidade científica.

- Estender o método proposto de detecção de semi-réplicas para possibilitar a recuperação de conjuntos de semi-réplicas na base de dados. Até então, apenas uma ou nenhuma semi-réplica é recuperada por imagem de consulta. Essa extensão pode ser implementada, sem complicações, considerando K-vizinhos mais próximos e retornando todas imagens com probabilidade acima de um limiar após um número mínimo pré-definido de amostras. Porém a base de dados deve ser trocada (já que a atual só possui uma semi-réplica por imagem de consulta).

- Estender o método proposto de detecção de semi-réplicas de forma a aproveitar o conhecimento do comportamento de distâncias entre vetores para diferentes tipos de transformações, calculando dinamicamente transformações prováveis entre imagens e retreinando o modelo estatístico.

- Estender o método proposto de detecção de semi-réplicas para palavras visuais baseando-se em [26], mas considerando o modelo Chi para distâncias entre correspondências.

Referências Bibliográficas

[1] Charu Aggarwal, Alexander Hinneburg, and Daniel Keim. On the surprising behavior of distance metrics in high dimensional space. *Database Theory (ICDT 2001)*, 1973:420–434, 2001.

[2] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc V. Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding (CVIU)*, 110(3):346–359, June 2008.

[3] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(4):509–522, 2001.

[4] Ondrej Chum, James Philbin, and Andrew Zisserman. Near duplicate image detection: min-hash and tf-idf weighting. In *Proceedings of the British Machine Vision Conference*, 2008.

[5] Ingemar J. Cox, Matt L. Miller, Thomas P. Minka, Thomas V. Papathomas, and Peter N. Yianilos. The bayesian image retrieval system, pichunter: Theory, implementation, and psychophysical experiments. *Transactions on image processing*, 9(1):20–37, 2000.

[6] S. Deb and Y. Zhang. An overview of content-based image retrieval techniques. In *Advanced Information Networking and Applications*, volume 1, pages 59–64. IEEE, 2004.

[7] George H. Dunteman. *Principal components analysis*, volume 69. SAGE, 1989.

[8] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Computer Vision and Pattern Recognition (CVPR05)*, volume 2, pages 524–531. IEEE, 2005.

[9] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 13(9):891–906, September 1991.

[10] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (2nd Edition)*. Prentice Hall, 2002.

[11] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Thirtieth annual ACM symposium on Theory of computing*, pages 604–613. ACM, 1998.

[12] E.T. Jaynes. *Probability Theory: The Logic of Science (Vol 1)*. Cambridge University Press, 2003.

[13] Yan Ke and Rahul Sukthankar. Efficient near-duplicate detection and sub-image retrieval. In *ACM Multimedia*, pages 869–876, 2004.

[14] Yan Ke and Rahul Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Conference on Computer Vision and Pattern Recognition (CVPR04)*, volume I, pages 511–517. IEEE, 2004.

[15] Yan Ke, Rahul Sukthankar, and Larry Huston. Efficient near-duplicate detection and sub-image retrieval. In *ACM Multimedia*, pages 869–876. ACM, 2004.

[16] C. T. Kelley. *Iterative Methods for Optimization*. SIAM Frontiers in Applied Mathematics, 1999.

[17] S. Lazebnik, C. Schmid, and J. Ponce. Sparse texture representation using affine-invariant neighborhoods. In *Conference on Computer Vision and Pattern Recognition*, volume 2, pages 319–324, 2003.

[18] Herwig Lejsek, Fridrik H. Ásmundsson, and Laurent Amsaleg Bjorn Thór Jónsson. Scalability of local image descriptors: A comparative study. In *14th international conference on Multimedia*, pages 589–598. ACM, 2006.

[19] Dimitri A. Lisin, Marwan A. Mattar, Matthew B. Blaschko, Erik G. Learned-Miller, and Mark C. Benfield. Combining local and global image features for object class recognition. In *Computer Vision and Pattern Recognition Workshop*, pages 47–54. IEEE, 2005.

[20] Wei Liu, Guangnan He, and Yubin Yang. Structural context: A new descriptor for object categorization. *Journal of Computer Science and Frontiers*, 4(4):304–311, April 2010.

[21] G. David Lowe. Object recognition from local scale-invariant features. In *Seventh International Conference on Computer Vision (ICCV99)*, volume II, pages 1150–1157. IEEE, 1999.

[22] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *British Machine Vision Conference*, volume 22, pages 761–767, 2002.

[23] Cláudia B. Medeiros. Grand research challenges in computer science in brazil. *Computer (Long Beach)*, 41(6):59–65, June 2008.

[24] Krystian Mikolajczyk and Cordelia Schmid. An affine invariant interest point detector. In *Seventh European Conference on Computer Vision (ECCV02)*, volume I, pages 128–142. Springer, 2002.

[25] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(10):1615–1630, October 2005.

[26] Andrej Mikulík, Michal Perdoch, Ondřej Chum, and Jiří Matas. Learning a fine vocabulary. In *11th European conference on computer vision (ECCV10)*, pages 1–14, 2010.

[27] Y. Rui, T. S. Huang, and S. Chang. Image retrieval: Current techniques, promising directions, and open issues. *Visual Communication and Image Representation*, 10(1):39–62, March 2002.

[28] Hans Sagan. *Space-Filling Curves*. Springer, 1 edition, 1994.

[29] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets. In *Seventh European Conference on Computer Vision (ECCV02)*, pages 414–431, 2002.

[30] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets, or how do i organize my holiday snaps? In *Seventh European Conference on Computer Vision (ECCV02)*, volume I, pages 414–431. Springer, 2002.

[31] J. Sivic and A. Zisserman. Video google: a text retrieval approach to object matching in videos. In *Ninth International Conference on Computer Vision (ICCV03)*, volume 2, pages 1470–1477. IEEE, 2003.

[32] G. Teodoro, D. Guedes, W. Meira, and R. Ferreira. Achieving multi-level parallelism in the filter-labeled stream programming model. In *International Conference on Parallel Processing (ICPP08)*, pages 287–294, 2008.

[33] M. Tico and P. Kuosmanen. Fingerprint matching using an orientation-based minutia descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 25(8):1009–1014, August 2003.

[34] Engin Tola, Vincent Lepetit, and Pascal Fua. A fast local descriptor for dense matching. In *Conference on Computer Vision and Pattern Recognition (CVPR08)*, pages 1–8. IEEE, 2008.

[35] T. Tuytelaars and L. Van Gool. Content-based image retrieval based on local affinely invariant regions. In *Conference on Visual Information Systems*, pages 493–500, 1999.

[36] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *11th British Machine Vision Conference*, pages 412–425, 2000.

[37] E. Valle, G. Teodoro, N. Mariano, R. S. Torres, and W. Meira. Adaptive parallel approximate similarity search for responsive multimedia retrieval. In *10th ACM Conference on Information and Knowledge Management (CIKM11)*. ACM, 2011.

[38] Eduardo Valle, Matthieu Cord, and Sylvie Philipp-Foliguet. 3-way-trees: A similarity search method for high-dimensional descriptor matching. In *International Conference on Image Processing (ICIP07)*, pages 173–176. IEEE, 2007.

[39] Eduardo Valle, Matthieu Cord, and Sylvie Philipp-Foliguet. Fast identification of visual documents using local descriptors. In *Proceeding of the eighth ACM symposium on Document engineering*, pages 173–176. ACM, 2008.

[40] Eduardo Valle, Matthieu Cord, and Sylvie Philipp-Foliguet. High-dimensional descriptor indexing for large multimedia databases. In *17th ACM Conference on Information and Knowledge Management (CIKM08)*, pages 739–748. ACM, 2008.

[41] Eduardo Valle, David Picard, and Matthieu Cord. Geometric consistency checking for local descriptor-based document retrieval. In *Symposium on Document Engineering (DocEng09)*, pages 135–138. ACM, 2009.

[42] Simon Winder, Gang Hua, and Matthew Brown. Picking the best daisy. In *Conference on Computer Vision and Pattern Recognition (CVPR09)*, pages 178–185. IEEE, 2009.

[43] Chunlei Yang, Jinye Peng, Xiaoyi Feng, and Jianping Fan. Speed up duplicate/near-duplicate image detection. In *International Conference on Internet Multimedia Computing and Service*, pages 95–98. ACM, 2010.

[44] Wan-Lei Zhao and Chong-Wah Ngo. Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection. *Image Processing, IEEE Transactions on*, 18(2):412–423, February 2009.