DETECÇÃO, RECONHECIMENTO E GEORREFERENCIAMENTO DE LOGOTIPOS

Letícia Aparecida de Moura Camelo[1], Francisco Assis da Silva[2], Ricardo Luís Barbosa[3], Mário Augusto Pazoti[2], Leandro Luiz de Almeida[2]

[1]Discente da Faculdade de Informática da UNOESTE. [2]Docente da Faculdade de Informática da UNOESTE. [3]Universidade Estadual Paulista (UNESP de Sorocaba/SP)

RESUMO

Sistemas Móveis de Mapeamento (SMM) Terrestres coletam dados para o mapeamento de vias e para alimentação de sistemas de informações geográficas. Eles integram sensores de imageamento digital, navegação e posicionamento embarcados, com isto fazem a coleta de dados para mapeamento de ruas e avenidas das cidades e possibilita o georreferenciamento dessas informações com rapidez, eficiência e economia de recursos. Uma das aplicações de um SMM é o georreferenciamento de pontos de interesse, como lojas, bancos etc. Porém, existem situações onde o trânsito de veículos não é permitido e o georreferenciamento tem que ser feito de forma manual. Este trabalho apresenta uma aplicação para a detecção e reconhecimento de logotipos, como o de grandes estabelecimentos comerciais, e georreferenciar os logotipos reconhecidos de maneira automática utilizando os recursos de um dispositivo móvel. Foram utilizados os algoritmos SIFT e RANSAC para a detecção e reconhecimento do logotipo em uma imagem obtida com a câmara do dispositivo móvel e para o georreferenciamento foi utilizado o GPS do próprio dispositivo móvel. Os resultados dos experimentos mostram que é possível o uso de um dispositivo móvel para realização do georreferenciamento de pontos de interesse onde o SMM não consegue trafegar.

Palavras-chave: reconhecimento de objetos; SIFT; georreferenciamento; mapeamento móvel.

LOGO DETECTION, RECOGNITION AND GEOREFERENCING

ABSTRACT

Mobile Mapping Systems (MMS) Land collect data for mapping routes and provide data for geographic information systems. They integrate digital imaging sensors, navigation and positioning embedded, thus they collect data for mapping streets and avenues of the cities and allow georeferencing this information with speed, efficiency and resource economy. One of the applications of a MMS is the georeferencing of interest points such as shops, banks etc. However, there are situations where vehicular traffic is not allowed and georeferencing has to be done manually. This paper presents an application for the detection and recognition of logos, such as large commercial and georeference logos recognized automatically using the resources of a mobile device. Were used SIFT and RANSAC algorithms for detection and recognition of the logo in an image taken from the camera of the mobile device and for the georeferencing was used GPS from mobile device. The results of the experiments show that it is possible to use a mobile device to perform the georeferencing of interest point where the MMS cannot to traffic.

Keywords: object recognition; SIFT; georeferencing; mobile mapping.

1. INTRODUÇÃO

Um dos métodos para o levantamento de dados geográficos com economia de recursos e rapidez são os Sistemas Móveis de Mapeamento (SMM). Estes sistemas desempenham um papel importante na coleta de dados para o mapeamento de ruas e na alimentação de sistemas de informações geográficas. O sistema de mapeamento móvel foi uma ideia de cientistas norte-americanos na década de 1980, que desenvolveram um meio de traçar mapas a partir de veículos terrestres. O SMM consiste em uma plataforma móvel que faz uso de sensores como INS (*Inertial Navigation System*), GPS (*Global Positioning System*), *laser scanner* e câmaras fotográficas e de vídeo para prover a coleta de informações sobre os atributos do espaço objeto a partir de imagens. Esses sensores são classificados em duas categorias: posicionamento e mapeamento (BARBOSA, 2006).

Os sensores de posicionamento podem ser externos ou internos. Sendo o sensor de posicionamento externo o GPS e o sensor interno o INS. Os sensores de mapeamento são classificados em passivos ou ativos. Os passivos são câmaras de vídeo ou fotográficas, e um sistema de varredura laser (*laser scanner*) o sensor de mapeamento ativo. Os sensores de posicionamento (GPS e INS) podem fornecer de maneira síncrona a orientação exterior dos sensores de mapeamento (câmaras de vídeo), de tal forma que os elementos de orientação exterior são conhecidos, e a reconstrução tridimensional pode ser efetuada diretamente utilizando pontos ou feições correspondentes.

Os SMM trafegam pelas vias das cidades ou estradas obtendo imagens e dados (coordenadas de posicionamento e ângulos de navegação) para capturar feições de interesse, possibilitando o armazenamento destes dados em um banco de dados e têm como uma de suas aplicações o georreferenciamento de pontos de interesses, como lojas, restaurantes, bancos etc. Esses pontos de interesse são detectados, reconhecidos e extraídos das imagens de maneiras manual, automática ou semiautomática e podem ser realizados em tempo real ou após o levantamento dos dados ter sido realizado.

Pelo fato de o SMM ser geralmente um veículo terrestre, existem situações onde não é possível o seu tráfego, como praças, calçadões, ruas intransitáveis etc. Com isso, o georreferenciamento para ser completo, deverá ser feito de forma manual, porém, o equipamento instalado no SMM, por ser de grande porte, não possibilita o georreferenciamento manual.

A alternativa proposta, neste trabalho, para detectar, reconhecer e georreferenciar pontos de interesse de forma manual sem o uso do SMM, é a utilização de um dispositivo móvel (iPhone/iPad), que é dotado de uma câmara fotográfica e sistema de posicionamento global (GPS). Para capturar as imagens e realizar todo o processamento de detecção e reconhecimento dos pontos de interesse, foi definida e implementada uma aplicação para que a atividade de georreferenciamento possa ser realizada.

A aplicação foi definida para conter um banco de *templates*, que precisa, em uma fase anterior ao do georreferenciamento, ser alimentado com as imagens dos templates que se deseja detectar e reconhecer.

Em uma fase anterior ao georreferenciamento deve-se realizar a detecção e reconhecimento dos logotipos, usando a correspondência entre duas imagens. A correspondência entre duas imagens é parte das aplicações da Visão Computacional. Para que essa correspondência aconteça é necessário encontrar primeiramente pontos em comum entre as imagens. Existem muitos trabalhos onde ocorre o emprego das mais diversas técnicas de processamento de imagens e reconhecimento de objetos, para que a identificação destes pontos aconteça de forma automática. Um exemplo é a técnica de descritores locais. Descritores são vetores de características de uma imagem ou de determinadas regiões de uma imagem e podem ser usados para se comparar regiões em imagens diferentes.

Neste trabalho, foi utilizado o algoritmo SIFT (LOWE, 2004) para a extração desses descritores, e com isso realizar a identificação de pontos chave entre a imagem obtida pela câmara do dispositivo móvel e as imagens de um conjunto de *templates* anteriormente cadastrados na aplicação. Para realizar a detecção e identificação do logotipo buscado é realizada a correspondência dos pontos chave de duas imagens, para com isso eliminar pontos que não condizem com o *template*. O algoritmo RANSAC (FISCHLER; BOLLES, 1981) foi usado neste processo.

Após esta seção introdutória, este trabalho está organizado conforme a estrutura que segue. Na Seção 2 são apresentados os trabalhos utilizados como base para a obtenção de conhecimentos para este trabalho; na Seção 3 são apresentados os algoritmos utilizados para a detecção e reconhecimento dos logotipos; na Seção 4 é apresentada a metodologia utilizada na definição de suas funcionalidades; na Seção 5 são apresentados os experimentos realizados para verificar a eficácia da aplicação

desenvolvida; por fim, a Seção 6 apresenta as conclusões e algumas sugestões para trabalhos futuros.

2. TRABALHOS RELACIONADOS

Na literatura, podem ser encontrados diversos trabalhos sobre georreferenciamento de pontos de interesse, detecção e reconhecimento de imagens, utilizando o algoritmo SIFT (LOWE, 2004). O algoritmo SIFT permite a detecção e extração de descritores que tem muitas propriedades que são almejáveis para correspondência de diferentes imagens de um objeto ou cena. Esses descritores são invariantes a mudanças de iluminação, ruído, rotação, escala e pequenas mudanças de ponto de vista.

São apresentados, na sequência, alguns trabalhos com aplicações de detecção, reconhecimento e georrefenciamento de imagens, que serviram de base para obtenção de conhecimentos para realização deste trabalho.

Amlacher e Paletta (2008) sugeriram o uso de imagens georreferenciadas para reconhecer um grupo de pontos turísticos por meio de serviços móveis de visão em cenários urbanos.

O trabalho de Baró et al. (2009a) propõe o reconhecimento de objetos em um banco de dados de imagens urbanas, e foi idealizado por causa do crescimento de conteúdo multimídia na internet. Em outro trabalho de Baró et al. (2009b) foi sugerida a construção de uma camada de conteúdo visual para reconhecimento de objetos em um banco de dados de imagens urbanas da cidade de Barcelona, com este sistema é possível de o usuário calcular de maneira automática a rota ideal para uma viagem e procurar locais como hotéis e restaurantes próximos a ele.

No trabalho de Maia (2010) é proposta uma metodologia para a utilização de mecanismos supervisionados de aprendizagem de máquina no contexto da detecção de pontos chave. Esta metodologia é inovadora, pois investiga a utilização de combinações não lineares de operadores diferenciais geométricos, que são obtidas de forma subentendida através do treino de um classificador não linear, usando *Support Vector Machine* (SVM). O principal objetivo dessa metodologia foi viabilizar a introdução de conhecimento externo sobre o conceito de pontos de interesse. Com esta metodologia observou-se que capacidade de generalização dos SVMs permite detectar pontos chave de forma estável, e a principal limitação desta metodologia consiste no tempo de processamento exigido para o treinamento e para a classificação.

Em Psyllos (2010) é apresentado um método para o reconhecimento de logotipos

de veículos. Inicialmente, o algoritmo SIFT (LOWE, 2004) foi aplicado em um conjunto de dados de treino para detectar a região de interesse em imagens e extrair as características discriminantes, que é utilizada como a assinatura de cada logotipo. Posteriormente, a estratégia de busca do vizinho mais próximo é construída. Finalmente, para reconhecer os logotipos na imagem de teste, algoritmo SIFT foi computado nas regiões de interesse da imagem. Resultados promissores foram obtidos em classificar milhares de logotipos nas imagens captadas por telefones celulares, em que a precisão do reconhecimento foi de 95%.

No trabalho de Romberg et al. (2011) foi proposto um *framework* para o reconhecimento de logotipos em imagens. Os autores definiram um método para codificar e indexar a disposição espacial relativa de características locais detectadas nas imagens do logotipo. Com base na análise dos recursos locais e da composição das estruturas espaciais básicas que pode derivar uma representação das regiões dos logotipos e minimizar as detecções de falsos positivos. Para a execução de testes foram usadas imagens de logotipos marcados manualmente e complementadas com imagens não logotipos. O conjunto de dados é composto por treinamento, validação e teste de um conjunto com 32 logotipos.

3. FUNDAMENTAÇÃO TEÓRICA

A identificação de pontos em comum entre duas imagens não é um trabalho simples. A primeira dificuldade está em encontrar pontos chave (*keypoints*) em uma das imagens e, em seguida, localizá-los na outra imagem. Neste trabalho, o algoritmo SIFT (LOWE, 2004) foi utilizado para encontrar *keypoints* entre imagem de entrada, que foi obtida através da câmara do dispositivo móvel, e nos *templates* (contendo imagens de logotipos populares). Depois de encontrados os *keypoints* entre as imagens, foi utilizado o algoritmo RANSAC para eliminar falsas correspondências entre os *keypoints*. Ambos os algoritmos são descritos nas seções seguintes.

3.1 SIFT

O SIFT (*Scale Invariant Feature Transform*) é um algoritmo de Visão Computacional proposto e publicado por David G. Lowe em 1999 (LOWE, 1999). Ele permite detecção e extração de um conjunto de vetores (descritores), onde cada um desses vetores é invariante a translação, rotação e escala e parcialmente invariante a iluminação e a projeções 3D. Estes descritores possuem muitas propriedades

que são necessárias para correspondência de diferentes imagens, eles são altamente distintos, sendo assim permite que uma determinada característica possa ser corretamente combinada.

O SIFT é composto de quatro etapas, onde as duas primeiras descrevem a parte do detector e as duas seguintes a formação do descritor. O detector SIFT é baseado em cálculos de diferença de Gaussianas e o descritor SIFT utiliza histogramas de gradientes orientados para descrever a vizinhança local dos pontos de interesse.

3.1.1 Etapas do SIFT

O algoritmo SIFT é executado em quatro principais etapas: Detecção do espaço de escala, detecção de extremos locais, atribuição de orientação e descrição dos *keypoints*. Sendo que as duas primeiras etapas fazem parte do detector e as duas últimas etapas são do descritor. A seguir, será descrito sucintamente o funcionamento de cada etapa que compõe o SIFT.

Detecção do espaço de escala: Nessa primeira etapa os *keypoints* são detectados por busca de características estáveis que identifica candidatos que são invariantes a escala e orientação. Isto é feito utilizando-se uma função chamada de espaço escala (WITKIN, 1983), no algoritmo SIFT é usada a função Gaussiana como sendo o núcleo (*kernel*) da função espaço escala. De maneira sucinta, são detectados extremos (máximos ou mínimos) em uma pirâmide da imagem convoluída com a função de diferença de filtros Gaussianos. O espaço escala de uma imagem é definido como uma função $L(x, y, \sigma)$, que é dada a partir da convolução de uma Gaussiana de escala variável $G(x, y, \sigma)$, com uma imagem de entrada $I(x, y)$, conforme a Equação 1:

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

onde * é a convolução de *x* em *y* com a Gaussiana

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

e σ representa o parâmetro que define o fator de suavização da imagem I.

Para detectar *keypoints* estáveis usando os extremos do espaço escala é utilizada a função DoG (*Difference of Gaussian*). A diferença de Gaussianas, representada por $D(x, y, \sigma)$, é calculada pela subtração de imagens filtradas em escalas próximas, separadas por uma constante multiplicadora *k*. A convolução de uma imagem com a função DoG é definida por:

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

Esta convolução é a diferença entre imagens suavizadas por um filtro Gaussiano em escalas *σ* e *kσ*. A função DoG consegue detectar variações de intensidade na

imagem, elimina detalhes indesejados e realça características fortes. Variando-se σ, é possível encontrar características em diferentes escalas. Um modo eficiente para a construção da DoG é por meio de uma pirâmide de imagens, mostrado na Figura 1. No algoritmo SIFT, primeiramente a imagem de entrada é redimensionada para o dobro de seu tamanho original, e isto é feito para se conseguir detectar mais pontos chave.

Para cada oitava no espaço escala, a imagem inicial sofre convoluções incrementais com filtros Gaussianos para produzir um conjunto de imagens separadas por um fator de escala constante *k* no espaço escala, gerando imagens suavizadas *L*, representadas na parte da esquerda da Figura 1. É denominado de oitava o conjunto de imagens *L* e DoG. A Figura 1 ilustra duas oitavas, estando à esquerda o conjunto de imagens suavizadas pelos filtros Gaussianos e à direita o conjunto de imagens resultantes da Diferença de Gaussianas.

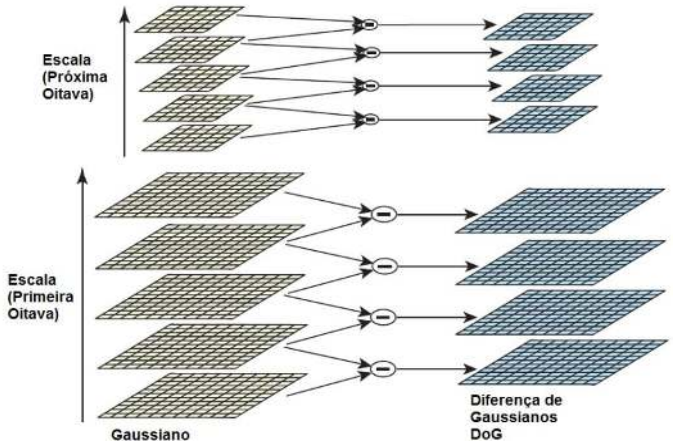


Figura 1. Representação do procedimento de obtenção das Diferenças de Gaussianas DoG para diversas oitavas de uma imagem (LOWE, 2004).

A imagem inicial é convoluída progressivamente com funções Gaussianas para produzir as outras *s* + 2 imagens utilizando um fator $k = 2^{1/s}$. São produzidas *s* + 3 imagens na pilha de imagens suavizadas com o filtro Gaussiano para cada oitava, de forma que a detecção de extremos cubra toda a oitava. A imagem é inicialmente filtrada por filtro Gaussiano com escala *s*. A partir de então, são geradas imagens que são progressivamente convoluídas, cada nova imagem é filtrada com escala *k* vezes a escala utilizada anteriormente. Imagens em escalas adjacentes são subtraídas para produzir imagens de diferença de Gaussiana, como mostrado na direita da Figura 1.

Após encontrar $D(x, y, \sigma)$, é feita a detecção de extremos em cada intervalo de

cada oitava. Os extremos são dados por valores de máximos ou mínimos locais de $D(x, y, \sigma)$, que são obtidos comparando cada ponto com seus oito vizinhos na sua escala, com os nove vizinhos na escala superior e os nove vizinhos da escala inferior, representados na Figura 2. Na figura, o ponto marcado com "X" é comparado com seus vizinhos marcados como "O". As 3 imagens DoG apresentadas na Figura 2 correspondem à diferença entre imagens adjacentes da pirâmide gaussiana.

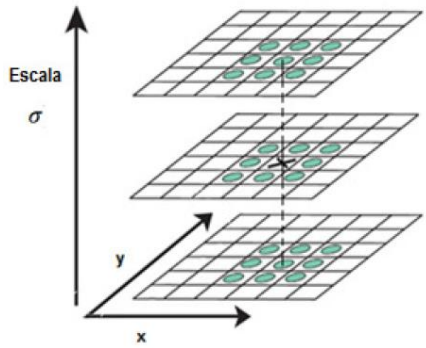


Figura 2. Detecção de extremos no espaço-escala (LOWE, 2004).

Localização de pontos chave: Todos os pontos detectados como extremos são possíveis candidatos a pontos chave e o próximo passo é calcular a localização exata destes pontos. Isto é feito basicamente ajustando uma função quadrática do ponto de amostragem local de modo a determinar uma localização interpolada do máximo.

A localização e a escala em que essa função alcança seu valor extremo nesta vizinhança passam a ser os novos valores de localização e escala do *keypoint*. Em Brown e Lowe (2002) demonstra-se que o deslocamento $\hat{x}$ da localização final do ponto chave em relação à posição x do extremo correspondente na DoG é dado por:

$$\hat{x} = \frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \quad (4)$$

Os coeficientes quadráticos são computados aproximando-se as derivadas através das diferenças entre pixels das imagens já filtradas. A localização do extremo, $\hat{x}$, é determinada ao se fazer a derivada de *D(x)* com relação à x, igualando o resultado a zero.

Definição da orientação dos pontos chave: É atribuída uma orientação a cada ponto chave, que será utilizada para construir descritores invariantes quanto à rotação. A orientação do ponto chave corresponde à direção predominante do gradiente em torno do ponto, onde a orientação é calculada no nível da pirâmide Gaussiana mais próxima à

escala onde se detectou o ponto chave. Computa-se o histograma de orientações dos gradientes desta imagem numa vizinhança em torno do ponto chave. A orientação do histograma tem 36 regiões cobrindo os 360 graus de orientações. Cada amostra na vizinhança do ponto chave é ponderada pela sua magnitude do gradiente e por uma janela Gaussiana circular com um σ que é 1.5 vezes o tamanho da escala σ do ponto chave.

Picos na orientação do histograma correspondem a direções predominantes para os gradientes locais. O maior pico no histograma é detectado e picos que correspondem a valores percentuais acima de 80% do maior pico são usados para se definir a orientação de cada ponto chave. Para se definir com maior precisão a orientação, uma parábola é ajustada entre os três valores do histograma mais próximos de cada pico, e então é interpolada a posição do pico para uma melhor precisão.

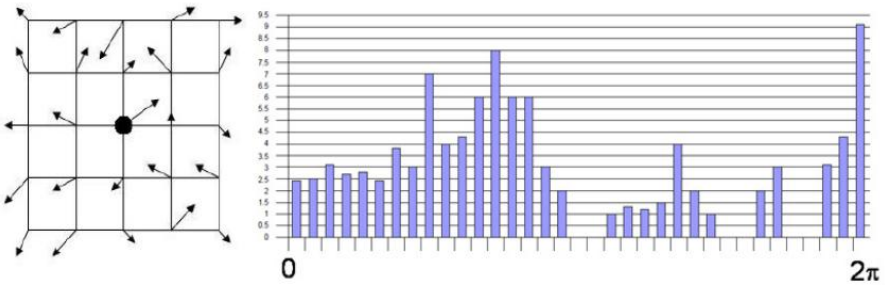


Figura 3. Histograma de orientação de um ponto chave (LOWE, 2004).

Construção do Descritor Local: Nesta etapa será atribuído a cada ponto chave um descritor invariante a iluminação e ponto de vista 3D. Este descritor do ponto chave é computado calculando-se inicialmente os gradientes que são amostrados em uma vizinhança em torno de cada ponto chave. Essa vizinhança é dividida em sub-regiões, onde os gradientes são representados pelas pequenas setas em cada amostra da localização, conforme a Figura 4.

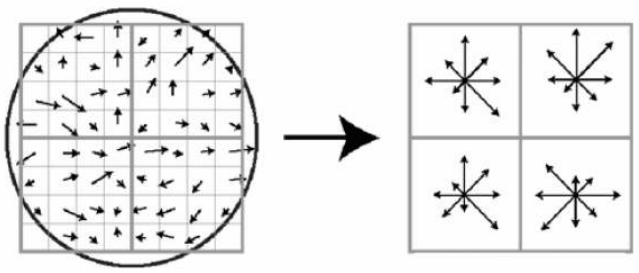


Figura 4. Gradientes da imagem e descritor de pontos chave (LOWE, 2004).

Para cada imagem, são construídos diversos descritores, cada um referente a um ponto chave. Tem-se como resultado um conjunto de descritores robustos que podem ser usados para fazer a correspondência da imagem em outra imagem.

3.1.2 Matching: Encontrando os Pontos em Comum

O conceito de *matching* é procurar os pontos correspondentes em cada uma das imagens. A comparação de pontos é fundamentada na semelhança dos descritores correspondentes.

Quando se trabalha com o algoritmo SIFT, pontos chave são detectados pelo algoritmo e representados por descritores. Os descritores são vetores de 128 posições que podem ser comparados utilizando a distância Euclidiana. Normalmente, os candidatos à melhor correspondência são pontos próximos, de maneira que o melhor candidato é o ponto que apresenta a menor distância Euclidiana.

Para evitar uma busca exaustiva, Lowe (2004) recomenda o uso de uma estrutura de dados *k-d tree*, que suporta uma busca binária balanceada para encontrar o vizinho mais próximo dos descritores. O algoritmo heurístico *Best-Bin-First* (BBF) é usado nesta busca, pois retorna o vizinho mais próximo, com alta probabilidade. O problema da correspondência é reduzido à busca do vizinho mais próximo. No entanto, alguns pontos instáveis são detectados durante o processo, levando a falsas correspondências. Para a eliminação desse problema, um método para comparar a menor distância com a segunda melhor distância é usado, selecionando somente correspondentes próximos de um limiar.

3.2 RANSAC

O algoritmo *RANdom SAmple Consensus* (RANSAC) foi proposto por Fischler e Bolles (1981) como um método para estimar os parâmetros de certo modelo a partir de um conjunto de dados com grandes quantidades de *outliers*[1].

O algoritmo é utilizado geralmente em Visão Computacional como, por exemplo, para resolver simultaneamente o problema de correspondência entre pontos de duas imagens e estimar a matriz fundamental relacionada ao par de imagens estéreo (MOREIRA, 2008). Uma de suas principais vantagens é ser um estimador robusto, ele pode estimar parâmetros com um alto grau de acerto. O RANSAC consegue evitar que *outliers* interfiram no modelo final.

[1] *Outliers:* pontos de dados que não se ajustam ao modelo correspondente ao objeto desejado estão fora de certa tolerância de erro.

3.2.1 Algoritmo

Para ser realizada a detecção e reconhecimento de objetos em imagens, adotando como base a correspondência dos pontos chave extraídos pelo algoritmo SIFT (LOWE, 2004), será necessário eliminar os pontos que não correspondem ou não se acertam com o modelo do logotipo desejado (*inliers*[2]), preservando apenas aqueles pontos chave que realmente condizem com o logotipo. Neste processo, o algoritmo RANSAC (FISCHLER; BOLLES, 1981) é um dos mais utilizado na literatura para o alcance de um resultado robusto.

O RANSAC foi desenvolvido com o intuito de evitar que *outliers* prejudiquem a computação do modelo final. Ao contrário das técnicas convencionais que utilizam muitos dados para obter uma solução inicial para então eliminar os *outliers*, o RANSAC utiliza apenas um conjunto com número mínimo e suficiente de pontos necessários para uma primeira estimativa, e procede aumentando o conjunto com pontos de dados consistentes. Uma amostra mínima depende do modelo a ser gerado. Geralmente, é o número *s* de dados que será utilizado para inicializar um modelo.

A partir do conjunto mínimo de pontos selecionados de maneira aleatória, o RANSAC mede a exatidão do modelo a ser estimado para todos os outros pontos, classificando-os em *inliers* ou *outliers*. Para tal, é definido um valor de limiar *t*, que é a distância máxima do modelo que um dado pode estar para ser considerado um *inlier*. O conjunto de *inliers* associado a um modelo é o conjunto de consenso do modelo. Assim, com várias iterações, o algoritmo é capaz de encontrar um modelo conciso para o problema em questão.

O tamanho do conjunto de consenso é o que determina o quão bom um modelo é. Um parâmetro *T* define o valor necessário que deve ser alcançado para um modelo ser considerado satisfatório. Durante a execução do RANSAC, quando for obtido mais de *T inliers*, o algoritmo finaliza, gerando novamente um novo modelo utilizando todos os *T inliers* obtidos, que será o modelo final encontrado pelo algoritmo. É preciso definir um número máximo *N* de iterações que o algoritmo irá realizar, para impedir um número infinito de iterações. Se um determinado modelo não alcançar *T inliers*, o algoritmo gera um novo modelo a partir de uma nova seleção de amostra mínima dos dados de entrada.

O algoritmo continua até conseguir um modelo aceitável, ou alcançar o número máximo de iterações. No caso de a execução do algoritmo finalizar por alcançar o número máximo de iterações, pode-se retornar o

[2] *Inliers*: pontos de dados que se ajustam com um determinado modelo desejado dentro de uma certa tolerância de erro.

conjunto com mais *inliers*, ou retornar que o algoritmo falhou.

Dado um modelo com parâmetros, deseja-se estimá-lo, para tal, os parâmetros podem ser estimados a partir de um número *S* de itens em um conjunto de dados conhecido. Os parâmetros que podem ser alterados de modo a determinar um bom resultado para o RANSAC são: *t*, que indica o limiar de distância que uma determinada entrada pode estar do modelo para ser considerado um *inlier*; *T*, que representa o número de *inliers* a ser alcançado para que um modelo possa ser considerado um bom modelo; e *N*, que representa o número máximo de iterações que devem ser realizadas.

O número de iterações *N* é escolhido grande o suficiente para garantir uma probabilidade *p*, de que pelo menos uma das amostras aleatórias do conjunto de ***s*** pontos não inclui um *outlier*. Considere *w* a representação da probabilidade que algum ponto de dado selecionado é um *inlier* e ε = 1 – *w* a probabilidade de observar um *outlier*. Pelo menos *N* iterações são requeridas, onde (HARTLEY; ZISSERMAN, 2003):

$$(1-w^s)^N = 1-p \qquad (5)$$

e assim o número de iterações necessárias é dado por:

$$N = \frac{\log(1-p)}{\log(1-(1-\varepsilon)^s)} \qquad (6)$$

4. METODOLOGIA APLICADA

Nesta seção é apresentada a metodologia aplicada neste trabalho para se realizar a detecção, reconhecimento dos logotipos e o georreferenciamento dos mesmos. Inicialmente são explicados os métodos utilizados, abordando resumidamente ao final o diagrama de atividade da metodologia utilizada.

Para a implementação da aplicação de detecção, reconhecimento e georreferenciamento de logotipos, foram desenvolvidas as seguintes etapas: criação de um banco de dados contendo os templates dos logotipos utilizados, as categorias nas quais estes templates se encaixam e os pontos de interesse encontrados pela aplicação; extração dos descritores da imagem de entrada e dos *templates*; reconhecimento dos logotipos e georreferenciamento dos mesmos.

A aplicação é composta por algumas funcionalidades que são relacionadas a seguir:

- Georreferenciamento: é nesta funcionalidade que, a partir da obtenção da imagem que contém o logotipo pela câmara do dispositivo, é realizado todo o

processo de detecção, reconhecimento e georreferenciamento dos logotipos. Os pontos de interesse resultantes deste processo são armazenados no banco de dados;

- Cadastro de Categorias: esta funcionalidade é utilizada para cadastrar o nome das categorias das quais os *templates* pertencerão;
- Cadastro de *Templates*: nesta funcionalidade são cadastrados os *templates* dos logotipos, composto pelo nome e da categoria da qual pertence;
- Pontos Recolhidos: nesta funcionalidade os pontos de interesse são recuperados e visualizados em um Mapa. Sendo que, cada ponto contém seu nome e categoria informados.

A Figura 5 apresenta um diagrama para demonstrar o funcionamento do sistema.

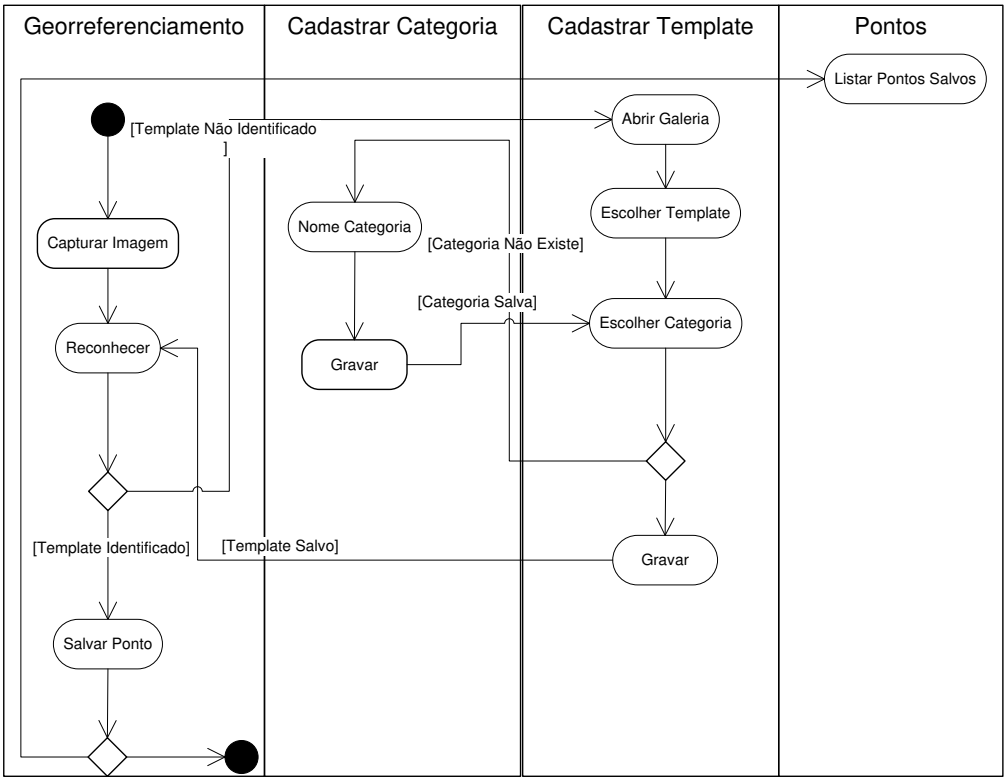


Figura 5. Diagrama da Atividade da metodologia utilizada para detecção, reconhecimento e georreferenciamento de logotipos.

Nas seguintes subseções são descritos o Banco de Dados definido e construído, a maneira de obtenção dos templates e como é realizada a detecção, reconhecimento e georreferenciamento dos logotipos.

4.1 Banco de dados

Foi construído um banco de dados utilizando o SQLite. O SQLite é uma biblioteca escrita na linguagem C que se comporta como um servidor de banco de dados e é a mais utilizada em aplicações iOS. Neste banco de dados são armazenados dados dos *templates* a serem comparados, as categorias das quais os *templates* pertencem e os dados dos georreferenciamentos realizados dos pontos de interesse, conforme ilustrado na Figura 6.

A seguir são apresentadas as tabelas, suas descrições e utilizações:

- "*categoria*": nesta tabela é armazenado o nome da categoria;
- "*templates*": nesta tabela são armazenados o nome do *template*, a imagem utilizada do *template* e a categoria da qual este template pertence;
- "*pontos*": nesta tabela são armazenados os dados principais da aplicação, tais como, o código *template* que foi reconhecido na imagem obtida pela câmara, o ponto de referência do local de onde a imagem foi obtida e os dados para o georreferenciamento: latitude, longitude e altitude.

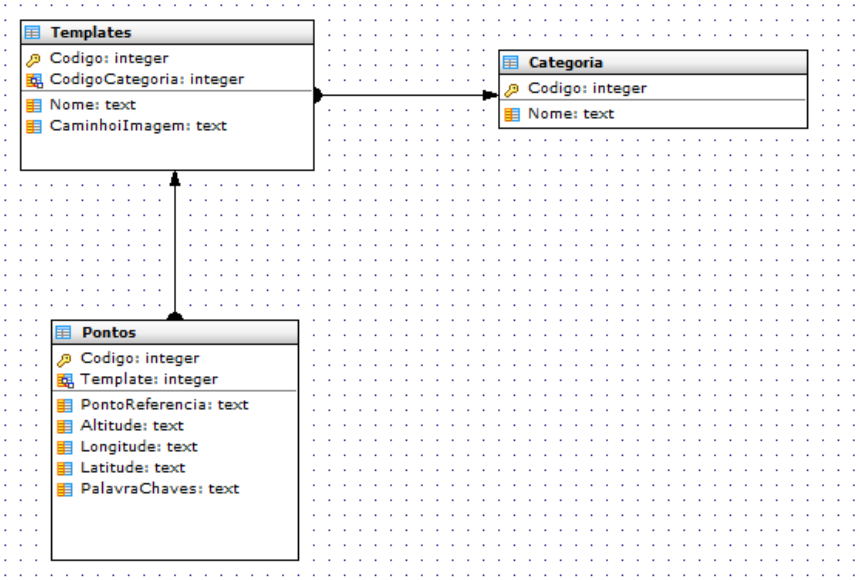


Figura 6. Modelo do Banco de Dados.

4.2 Templates e Categorias

Antes da fase de detecção e reconhecimento, são definidos os *templates* que podem ser obtidos com a própria câmara do dispositivo móvel e as categorias as quais estes *templates* pertencem. Para o cadastro das categorias faz-se necessário apenas informar o nome da mesma. No cadastro de *templates* os seguintes dados são gravados no banco de dados para o uso posterior do algoritmo de reconhecimento: nome do *template*, a imagem do logotipo e a categoria a qual o *template* pertence.

4.3 Detecção, reconhecimento e georreferenciamento

Neste trabalho foi utilizado para a extração de descritores o algoritmo SIFT (LOWE, 2004) que é razoavelmente invariante à mudança de iluminação, ruído, rotação, escala e pequenas mudanças no ponto de vista. A seção 3.1 descreve com detalhes todos os estágios do algoritmo SIFT.

Os descritores extraídos com o algoritmo SIFT da imagem obtida pela câmara do dispositivo são correspondidos (*matching*) com as imagens dos *templates*, que já foram anteriormente armazenados em um banco de dados. O *matching* entre os descritores extraídos da imagem obtida pela câmara e dos *templates* armazenados deve ser o de menor distância Euclidiana entre o vetor extraído na sequência e o vetor do *template.* A partir dos pontos que são correspondentes entre as duas imagens ocorre a extração dos pontos que condizem com o modelo do template do logotipo (*inliers*) e a eliminação dos pontos que não condizem (*outliers*).

Após a detecção e reconhecimento do *template* na imagem obtida é feito o georreferenciamento do ponto de interesse. Este georrefenciamento é realizado de maneira automática usando o GPS do dispositivo móvel, somente será necessário se aproximar do local onde foi obtida a imagem pela câmara para que o georreferenciamento possa ser mais preciso.

5. EXPERIMENTOS E RESULTADOS

Neste trabalho foram realizados experimentos para verificar a eficiência da aplicação desenvolvida, se os logotipos estão sendo devidamente detectados, reconhecidos e georreferenciados.

Os testes foram realizados com imagens de fachadas que possuem os *templates* utilizados neste experimento. Na Figura 7, estão relacionados os *templates* utilizados para realização dos testes: (1) Banco do Brasil; (2) Universidade do Oeste Paulista – Unoeste; (3) Banco Itaú; (4) Supermercado Carrefour; (5) Casas Bahia; (6) Ponto Frio.



Figura 7. Logotipos utilizados.

Foi aplicada nas imagens escolhidas a metodologia proposta neste trabalho. As imagens de entrada foram capturadas com a câmara do dispositivo móvel, sendo que todas as imagens possuíam os *templates* cadastrados. Com a metodologia adotada neste trabalho, houve o reconhecimento total dos *templates*, conforme pode ser observado na Figura 8, sendo as demarcações em cor verde os *templates* detectados e reconhecidos.

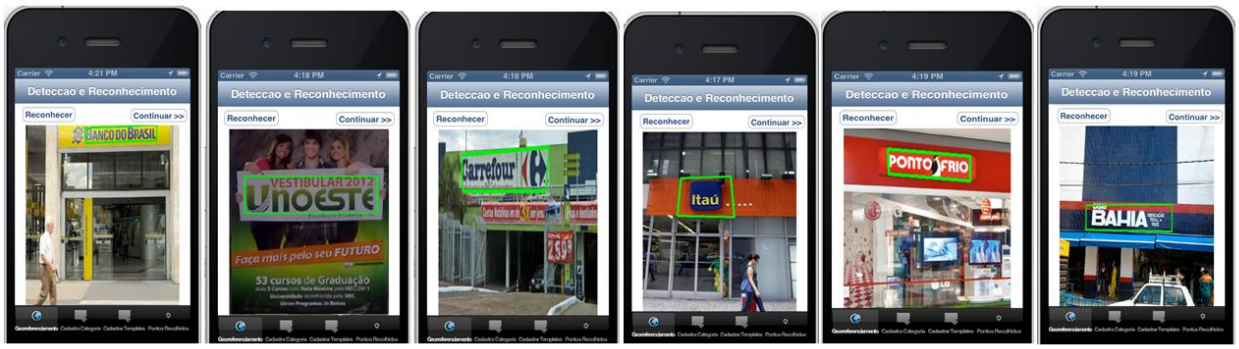


Figura 8. Resultados obtidos com o processo de detecção e reconhecimento.

A Figura 9 ilustra um mapa contendo o local onde um dos logotipos testados, que foi detectado, reconhecido e georreferenciado pela aplicação desenvolvida.

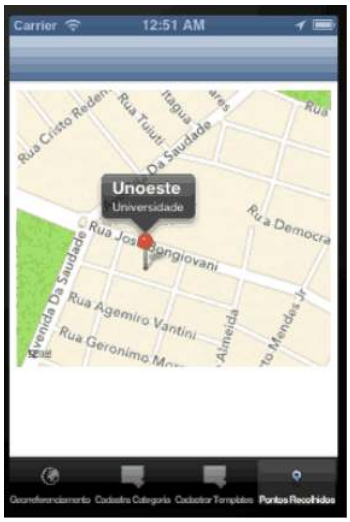


Figura 9. Ponto de interesse georreferenciado.

6. CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho apresentou a proposta de uma aplicação em um dispositivo móvel para realizar a detecção, o reconhecimento e o georreferenciamento de logotipos, usando imagens obtidas com a câmara do próprio dispositivo.

Os resultados obtidos com a aplicação proposta demonstra que é possível o uso de equipamentos de baixo custo, como o dispositivo móvel, para realização do georreferenciamento de pontos de interesse onde o SMM não consegue trafegar.

Em trabalhos futuros esta aplicação poderá ser utilizada para a alimentação de sistemas de localização para fins turísticos, e também para alimentação bases de dados para fins de localização de estabelecimentos comerciais como: bancos, universidades, bares, cafés, restaurantes, lojas etc.

REFERÊNCIAS

AMLACHER, K.; PALETTA, L. Geo-indexed object recognition for mobile vision tasks. In: INTERNATIONAL CONFERENCE ON HUMAN COMPUTER INTERACTION WITH MOBILE DEVICES AND SERVICES, 10. **Proceedings…** 2008. p. 371-374.

BARBOSA, R.L. **Caminhamento fotogramétrico utilizando o fluxo óptico filtrado.** 2006. 118 f. Tese (Doutorado em Ciências Cartográficas) – Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, Presidente Prudente, 2006.

BARÓ, X. et al. Generic **Object recognition in urban image databases**. In: INTERNATIONAL CONFERENCE OF THE CATALAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE, 12. **Proceedings…** 2009a. p. 27-34. http://dx.doi.org/10.1109/ICME.2009.5202828

BARÓ, X. et al. Visual content layer for scalable object recognition in urban image databases. In: 2009 IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO. **Proceedings…** 2009b. p. 1616-1619.

BROWN, M.; LOWE, D.G. Invariant features from interest point groups. In: BRITISH

MACINHE VISION CONFERENCE. **Proceedings...** Cardiff, Wales, 2002, p. 656-665.

FISCHLER, M.A.; BOLLES, R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. **Communications of the ACM**, v.24, p.381-395, 1981. http://dx.doi.org/10.1145/358669.358692

HARTLEY, R.; ZISSERMAN, A. **Multiple view geometry in computer vision**. 2. ed. Cambridge: University Press, 2003.

LOWE, D.G. Object recognition from local scale invariant feature. **International Journal of Computer Vision**, Corfu, p. 1159-1157, 1999.

LOWE, D.G. Distinctive image features from scale-invariant keypoints, International **Journal of Computer Vision**, v. 60, n.2, p. 91-110, 2004. http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94

MAIA, J.G.R. **Detecção e reconhecimento de objetos usando descritores locais.** 2010. 127f. Tese (Doutorado em Ciência da Computação) - Universidade Federal do Ceará.

MOREIRA, G. **Relatório sobre os resultados obtidos através do uso dos algoritmos SIFT e RANSAC para reconstrução de um objeto a partir de uma nuvem de pontos.** PUC-Rio, Departamento de Informática. 2008. Disponível em: <http://www.tecgraf.puc-rio.br/~mgattass/ra/trb08/GustavoMoreira/rel_SIFT_RANSAC.pdf>

PSYLLOS, A.P. Vehicle logo recognition using a SIFT-Based enhanced matching scheme. **IEEE Intelligent Transportation Systems Society**, v.11, n.2, p.322-328, 2010. 10.1109/TITS.2010.2042714

ROMBERG, S. et al. **Scalable logo recognition in real-world images**. Alemanha: Universidade de Augsburg: Departamento de Informática, 2011.

WITKIN, A.P. Scale-space filtering. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE. **Proceedings...** Karlsruhe, Germany. p.1019-1022, 1983.