

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Um ambiente para anotação de localização e eventos em coleções de fotografias

Davi Oliveira Serrano de Andrade

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande – Campus I como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Sistemas de Informação e Banco de Dados

Prof. Dr. Cláudio de Souza Baptista
(Orientador)

Campina Grande, Paraíba, Brasil

© Davi Oliveira Serrano de Andrade, 19/03/2015

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

A553a Andrade, Davi Oliveira Serrano de.

Um ambiente para anotação de localização e eventos em coleções de fotografias / Davi Oliveira Serrano de Andrade. – Campina Grande, 2015.
100 f.: il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2015.

"Orientação: Prof. Dr. Cláudio de Souza Baptista".

Referências.

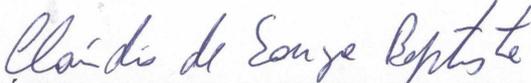
1. Multimídia. 2. Anotação de Localização. 3. Anotação de Eventos.
4. Fotografia. I. Baptista, Cláudio de Souza. II. Título.

CDU 004.032.6(043)

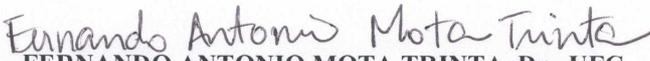
**"UM AMBIENTE PARA ANOTAÇÃO DE LOCALIZAÇÃO E EVENTOS EM COLEÇÕES
DE FOTOGRAFIAS"**

DAVI OLIVEIRA SERRANO DE ANDRADE

DISSERTAÇÃO APROVADA EM 19/03/2015


CLÁUDIO DE SOUZA BAPTISTA, PhD., UFCG
Orientador(a)


LEANDRO BALBY MARINHO, Dr., UFCG
Examinador(a)


FERNANDO ANTONIO MOTA TRINTA, Dr., UFC
Examinador(a)

CAMPINA GRANDE - PB

Resumo

Em razão do elevado número de fotografias gerado atualmente, técnicas para organizar, buscar e recuperar tais imagens são fundamentais. Organizar uma coleção de fotografias com milhares de imagens não é um trabalho simples. Além disso, associar dados de redes sociais com fotografias é ainda mais trabalhoso. Fotografias pessoais são comumente anotadas tomando-se como referência as seguintes perguntas: "Quem? Onde? Quando?". Considerando-se as informações importantes na recuperação de fotografias, este trabalho centra esforços nas questões "Onde?" e "Quando?". Com essas duas perguntas em mente, o foco está voltado para a localização ("Onde?") e os eventos da fotografia ("Onde?" e "Quando?"). O objetivo geral consiste em propor um ambiente para anotação de localização e eventos sociais. Essa anotação é auxiliada pelas técnicas de propagação de localização e de detecção de eventos sociais propostas neste trabalho. Os resultados dos experimentos com técnicas de propagação de localização indicam que a escolha dessa técnica deve ocorrer conforme o comportamento de cada usuário do sistema. Por isso, além das técnicas de propagação, propõe-se, neste trabalho, uma seleção automática de técnica de propagação de localização. Os experimentos realizados para validar a técnica de detecção de eventos sociais apresentaram bons resultados e a referida técnica, além de realizar a detecção de eventos, também pode ser usada para agrupar fotografias pertencentes a um mesmo evento. Por fim, este trabalho apresenta um protótipo de ferramenta *web* para unir a anotação de localização com a anotação de eventos.

Abstract

Due to the large number of pictures that is currently generated, it is very important to have techniques to organize, search and retrieve such images. Organize a collection of photos with thousands of images is not a simple job, and to associate data from social networks with the photographs is even more laborious. Personal photographs are commonly organized by reference to the following questions: "Who? Where? When?". Considering the three important questions, this work focuses efforts on the questions "Where?" and "When?". With these two questions in mind, the focus will be on location ("Where?") and events ("Where?" and "When?"). The overall objective is to offer an environment for annotating location and social events. This annotation is aided by the techniques proposed in this work. The experiments with location propagation techniques indicate that the choice of the propagation technique should happen considering the behavior of each user. Therefore, in addition to propagation techniques, is proposed, in this work, an automatic selection of location propagation techniques. The experiments performed to validate the social event detection technique presented good results and, in addition to performing event detection, can also be used to group photographs belonging to the same event. Finally, this work presents a tool to unite the annotation location with the annotation of events.

Agradecimentos

Ao único Deus do universo, aquele que é digno de toda honra e glória, que enviou seu filho Jesus para morrer na cruz para nos salvar. Ele que, mesmo eu sendo falho, me dá razão para viver. Ao Deus que me ama incondicionalmente, a quem buscarei demonstrar minha gratidão com todas as minhas forças.

Aos meus pais, por servirem de inspiração e exemplo durante toda a minha vida, por terem sempre acreditado no meu potencial, por terem me passado princípios éticos e morais, e, principalmente, por terem me ensinado que esses princípios possuem um valor imensurável.

A Moema, minha namorada, que Deus colocou em minha vida para compartilhar um amor puro e verdadeiro. Obrigado por me fazer completo, por estar sempre ao meu lado como amiga, companheira e cúmplice.

Ao orientador, amigo e exemplo, Prof. Dr. Cláudio de Souza Baptista, pela orientação exemplar, por sempre exigir um trabalho de qualidade e, sobretudo, por ter sido um exemplo de ser humano.

À minha família, formada pela “Serranada” e pela “Grande Família Oliveira”, por cuidar de mim e sempre ser exemplo de união.

Ao meu irmão, Daniel, pela companhia durante todo o trabalho.

Ao amigo Hugo Feitosa, com quem realizei bons trabalhos, pelo acompanhamento e pelos conselhos. Aos colegas do LSI, por todos os momentos de descontração e por tornarem o ambiente de trabalho um lugar bom.

Aos amigos Fábio Leal e Felipe Neves, pelo eterno companheirismo e por me receberem em suas casas quando precisei.

Aos amigos Demontiê, Izabela, João Paulo, Laerte, Raíssa, Raquel e Savyo, pelas saídas e pelos momentos de diversão.

Aos amigos Arthur Marques e Bruno Dias, pelos almoços inesquecíveis.

Aos amigos George Henrique, Ana Marília, Diogo Schulz, Raissa Borges, Mateus Cabral e Clarissa Guedes, que me propiciaram momentos de muita alegria.

À UFCG e ao LSI, por toda minha formação acadêmica.

À CAPES e ao CNPQ, pelo apoio financeiro.

Conteúdo

Capítulo 1 – Introdução	1
1.1 Problemática.....	2
1.1.1 Localização	2
1.1.2 Eventos	3
1.2 Objetivo.....	4
1.2.1 Objetivos específicos.....	5
1.3 Contribuições	6
1.3.1 Publicações	6
1.4 Estrutura da dissertação.....	7
Capítulo 2 – Fundamentação teórica	8
2.1. Anotação de fotografias	8
2.1.1 Contexto.....	9
2.1.2 Localização	10
2.1.3 Eventos sociais	11
2.1.4 Métricas	12
2.2. Algoritmos.....	13
2.2.1 Regressão linear múltipla	13
2.2.2 Fatoração de matriz	15
2.2.3 Máquina de vetores de suporte (SVM).....	16
2.2.4 Indexação TF-IDF	17
2.2.5 Similaridade de cosseno com TF-IDF	18
2.2.6 Agrupamento DBSCAN	19
2.3. Considerações finais.....	20
Capítulo 3 – Trabalhos relacionados	21
3.1 Localização geográfica.....	21
3.2 Eventos	22
3.3 Comparativo dos trabalhos relacionados.....	29
3.4 Considerações finais.....	32
Capítulo 4 – Anotação de localização	33
4.1 Propagação de localização	36
4.1.1 Técnica de eventos compartilhados	37
4.1.2 Técnica de trajetória	37

4.1.3 Exemplos	38
4.1.4 Técnicas reproduzidas	40
4.1.5 Validação	42
4.1.6 Resultados.....	46
4.2 Seleção automática de técnica de propagação.....	49
4.2.1 Modelo de regressão linear múltipla	54
4.2.2 Validação	56
4.2.3 Resultados.....	59
4.3 Considerações finais.....	61
Capítulo 5 – Anotação de eventos	63
5.1 Agrupamento de fotografias.....	64
5.2 Propagação de eventos	67
5.3 Validação.....	70
5.3.1 Base de dados	70
5.3.2 Validação do agrupamento de fotografias	70
5.3.3 Validação da propagação de eventos.....	72
5.4 Resultados	74
5.5 Considerações finais.....	79
Capítulo 6 – Protótipo de ferramenta.....	80
6.1 Minimização da quantidade de fotografias a serem anotadas	82
6.2 A arquitetura da ferramenta PG++	83
6.3 Anotação de localização.....	85
6.3.1 Extração de informações das fotos	86
6.3.2 Organização automática de álbuns	86
6.4 Anotação de eventos.....	88
6.4 Considerações finais.....	89
Capítulo 7 – Conclusões e trabalhos futuros	90
7.1 Contribuições	92
7.2 Trabalhos futuros.....	93
Referências bibliográficas.....	94

Lista de abreviaturas

UFCG	Universidade Federal de Campina Grande
LSI	Laboratório de Sistemas de Informação
EXIF	<i>Exchangeable image file format</i>
GPS	<i>Global Positioning System</i>
A-GPS	<i>Assisted Global Positioning System</i>
IBERAMIA	Ibero-American Conference on Artificial Intelligence
ICEIS	International Conference on Enterprise Information Systems
WFA	<i>Workshop de Ferramentas e Aplicações</i>
W3C	<i>World Wide Web Consortium</i>
TF-IDF	<i>Term Frequency–Inverse Document Frequency</i>
DBSCAN	<i>Density-based spatial clustering of applications with noise</i>
SVM	<i>Support Vector Machines</i>
CrEve	<i>Collaborative Event Annotation Framework</i>
GESODECK	<i>Geo-social event detection and tracking system</i>
URL	<i>Uniform Resource Locator</i>
k-NN	<i>k-Nearest Neighbors</i>
SED	<i>Social Event Detection</i>
HTML	<i>HyperText Markup Language</i>
JSON	<i>JavaScript Object Notation</i>
SQL	<i>Structured Query Language</i>
JPEG	<i>Joint Photographic Experts Group</i>
TIFF	<i>Tagged Image File Format</i>

Lista de figuras

Figura 1.1. Tipos de Eventos	4
Figura 2.1. Exemplo de propagação com limiar espacial	11
Figura 2.2. Regressão linear simples	13
Figura 2.3. Regressão linear múltipla	14
Figura 2.4. Estimção de valores com fatoração de matriz	16
Figura 2.5. Hiperplano encontrado com SVM.....	17
Figura 2.6. Exemplo de agrupamento com DBScan.....	19
Figura 4.1. Exemplo com algoritmo baseado na trajetória	39
Figura 4.2. Fotografias dos amigos no evento	40
Figura 4.3. Acurácia das Técnicas	47
Figura 4.4. Cobertura das Técnicas.....	47
Figura 4.5. Precisão das Técnicas	47
Figura 4.6. Comportamento por Usuário	49
Figura 4.7. Exemplo de tabela de um usuário aleatório.....	53
Figura 4.8. Modelo da seleção automática proposta neste trabalho	55
Figura 5.1. Grupos temporais e grupos espaciais	66
Figura 5.2. Anotação de evento em fotografia com localização.....	68
Figura 5.3. Anotação de eventos com TF-IDF	68
Figura 5.4. Fluxograma do algoritmo Muído.....	69
Figura 5.5. Melhores configurações de agrupamento em Pegado	74
Figura 5.6. Precisão da anotação automática em fotografias com localização.....	75
Figura 5.7. Cobertura da anotação automática em fotografias com localização	76
Figura 5.8. Medida F da anotação automática em fotografias com localização	77
Figura 5.9. Cobertura da anotação automática em fotografias sem localização.....	78
Figura 5.10. Medida F da anotação automática em fotografias sem localização	79
Figura 6.1. Tela de <i>login</i> e cadastro da ferramenta.....	81
Figura 6.2. Tela inicial da ferramenta.....	81
Figura 6.3. Tela de anotação de localização dos grupos.....	83
Figura 6.4. Arquitetura da ferramenta PG++	84
Figura 6.5. Minimização da quantidade de fotografias anotadas manualmente.....	86
Figura 6.6. Criação dos geoálbuns.....	87
Figura 6.7. Interface de apresentação dos geoálbuns.....	87
Figura 6.8. Anotação de Eventos	88

Lista de Tabelas

Tabela 3.1. Estudo comparativo dos trabalhos estudados.....	31
Tabela 4.1. Dicionário de dados referentes à tabela de metadados.	51
Tabela 4.2. Confirmação das hipóteses.....	60
Tabela 4.3. Diferença média entre a seleção automática proposta e a melhor técnica do cenário.....	61

Lista de Algoritmos

Algoritmo 4.1. Coleta dos resultados.....	45
Algoritmo 5.1. Validação do agrupamento de fotos	66
Algoritmo 5.2. Validação do agrupamento de fotos	71
Algoritmo 5.3. Validação da propagação de eventos	73
Algoritmo 6.1. Agrupamento das fotografias para minimização da anotação manual.	82

Capítulo 1 – Introdução

Os avanços tecnológicos no mercado de fotografias nos últimos anos tornaram muito simples o processo de captura de imagens. Esses avanços, em conjunto com a diminuição no custo de armazenamento, levaram a um grande uso de dispositivos eletrônicos, tais como câmeras digitais, *smartphones* e *tablets*. Em decorrência dessa utilização de dispositivos, as pessoas geram considerável número de arquivos de multimídia, dificultando a tarefa de anotar e catalogar todo esse material.

Organizar manualmente uma coleção com milhares de fotografias tiradas durante uma viagem de férias, por exemplo, é uma tarefa bastante onerosa. Essa realidade desencadeou a discussão de como as fotografias são organizadas (RODDEN e WOOD, 2003). Várias abordagens têm sido sugeridas para organização automática de fotografias com o objetivo de minimizar o trabalho do usuário (FIGUEIRÊDO et al., 2012; LACERDA et al., 2008; NAAMAN e NAIR, 2008). Paralelamente à criação de tais sistemas de organização, também surgiu a criação dos metadados das fotografias para melhorar a organização de coleções de fotografias.

Alguns exemplos dessas informações são data, hora, localização geográfica e *tags*. Com o significativo aumento de fotografias das coleções, gerenciar e organizar todas elas manualmente torna-se inviável, pois a quantidade de fotografias e, conseqüentemente, de metadados é muito grande. Imagine-se o árduo trabalho de organizar manualmente uma coleção com mil fotografias em relação às *tags*.

Fotografias pessoais são comumente anotadas tomando como referência as seguintes perguntas: "Quem? Quando? Onde?" (NAAMAN et al., 2004). "Quem?" representa as pessoas presentes em uma determinada fotografia. "Quando?" indica o instante em que a fotografia foi capturada. "Onde?" retrata o lugar onde a fotografia foi tirada. Serve como exemplo uma fotografia de três amigos, na virada do ano de 2015, em Copacabana. Os três amigos que aparecem na fotografia são a resposta à pergunta "Quem?", o instante de tempo 00:00:01 do dia 01/01/2015 é a resposta à pergunta "Quando?" e Copacabana/Rio de Janeiro é a resposta à pergunta "Onde?". A localização também pode ser representada como uma coordenada geográfica ou descrição do lugar.

O lugar onde a fotografia foi tirada é uma das primeiras informações que as

pessoas lembram quando querem recuperar aquela imagem. Isso significa que a localização geográfica da câmera naquele momento é muito importante nesse processo de organização. Além da localização, o evento em que a fotografia se dá também é uma informação importante. Segundo Cooper et al. (2005), "*Eventos estão naturalmente associados com lugares e tempos específicos, como uma festa de aniversário de uma criança, ou umas férias*". Ainda no exemplo supracitado, o evento associado à fotografia seria “Réveillon em Copacabana”, associado ao lugar “Copacabana” e ao intervalo de tempo específico da festa da virada do ano.

1.1 Problemática

Em virtude do significativo número de fotografias gerado atualmente, técnicas para organizar, buscar e recuperar tais imagens são fundamentais. Organizar uma coleção de fotografias com milhares de imagens não é um trabalho simples. Além disso, associar dados de redes sociais com as fotografias é ainda mais trabalhoso. Várias abordagens foram criadas para solucionar esse problema, como o PhotoGeo (FIGUEIRÊDO et al., 2012; LACERDA et al., 2008), que é uma ferramenta criada na academia e apresenta funcionalidades, como anotação de pessoas, eventos pessoais e localização.

Considerando-se as informações importantes na recuperação de fotografias, este trabalho centra esforços nas questões “Onde?” e “Quando?”. Com essas duas perguntas em mente, o escopo do problema de anotação em fotografias pessoais está limitado à localização e ao evento da fotografia. A localização é a resposta à pergunta “Onde?” e os eventos estão diretamente ligados a um lugar (“Onde?”) e a um tempo específico (“Quando?”). Para anotação de localização e eventos, as técnicas existentes são sensíveis aos dados de treinamento. O problema de anotação é apresentado considerando-se o contexto de localização de fotografias, seguido do contexto de eventos em coleções de fotografias.

1.1.1 Localização

Com o passar do tempo, a integração do *Global Position System* (GPS) nos *smartphones*, *tablets* e câmeras digitais torna-se cada vez mais presente, permitindo o armazenamento automático da localização geográfica nos metadados das fotografias. Receptores GPS, no entanto, não funcionam bem em locais fechados, podendo não

gerar informação alguma. Vale ressaltar, ainda, que são necessários alguns instantes para que o receptor receba o sinal, indicando a localização da fotografia. Se o sinal não é capturado, a câmera usará a localização da última fotografia (o que possivelmente será uma informação errada) ou não fará o georreferenciamento devido. Alguns *smartphones* usam a tecnologia A-GPS para minimizar esses problemas, porém as informações geradas podem ser bem distantes da localização real onde a fotografia foi tirada.

Para que o trabalho de organizar fotografias do usuário seja otimizado, é necessário ter a informação geográfica. O uso de GPS facilita a coleta da localização das fotografias, porém as falhas do próprio receptor causam um trabalho adicional para o usuário. Corrigir manualmente as falhas causadas pelo receptor GPS é muito difícil, daí a necessidade de trabalhos na área de propagação e/ou predição de localização para melhorar os resultados de recuperação de multimídias (HAYS e EFROS, 2008; JOSHI et al., 2012; LACERDA et al., 2013).

Além das limitações técnicas, o comportamento do usuário pode influenciar. Existem usuários que se deslocam bastante, logo suas fotografias pertencerão a muitos lugares distintos que estarão associados ao seu deslocamento individual. Dessa forma, a trajetória do usuário deve ser levada em conta para encontrar a localização das fotografias. Há usuários que costumam frequentar muitas festas, fazendo com que a localização de suas fotografias esteja diretamente ligada aos eventos sociais frequentados. Nesses eventos, outras pessoas capturam fotografias e as informações atreladas a essas podem ser usadas para encontrar a localização de fotografias no mesmo evento. Como existe uma diversidade no comportamento dos usuários e essa diferença de comportamento pode interferir na anotação automática de localização, a anotação de localização não consegue adaptar-se à realidade de cada um e termina prejudicando uma parte dos usuários.

1.1.2 Eventos

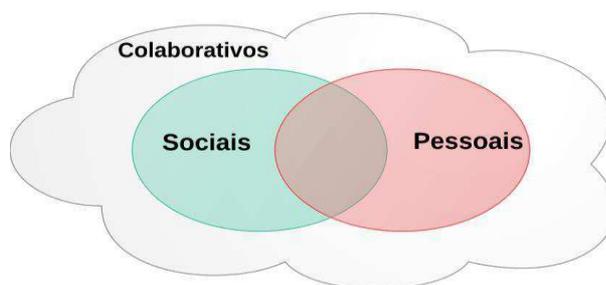
A organização de fotografias está se aprimorando com o passar do tempo e várias abordagens estão sendo propostas para organizar as coleções de fotografias de usuários. De acordo com Lim et al. (2003), eles preferem organizar suas fotografias em eixos semânticos, como eventos, pessoas, dentre outros. Cooper et al. (2005) afirma que "*Os usuários constantemente organizam suas fotografias em termos de 'eventos' tanto para busca e recuperação, quanto para compartilhar fotografias selecionadas*

com outros". Dessa forma, a organização de fotografias em eventos é fundamental para se chegar a uma automatização adequada.

Como organizar as fotografias manualmente levando em conta os eventos associados a elas é difícil, é possível que o usuário escolha uma opção menos trabalhosa, mas que não auxilie futuras buscas e recuperações. As pesquisas na área seguem abordagens diferentes na organização de fotografias em eventos.

Trabalhos como os de Becker et al. (2010) e Brenner et al. (2012) focam na linha de **eventos sociais**, que são os eventos externos, ou seja, não relacionados com a vida pessoal do usuário que tem a fotografia. Outros trabalhos como Cooper et al. e Figueirêdo et al. (2012) abordam os **eventos pessoais**, que são os relacionados com a vida pessoal do usuário. Além desses tipos de eventos, ainda existem trabalhos, como o de Zigkolis et al. (2012), que focam a ideia de **eventos colaborativos**, que são eventos criados pela colaboração dos usuários da aplicação.

Figura 1.1. Tipos de Eventos



Rock in Rio e Copa do Mundo são exemplos de eventos sociais, assim como formatura, viagem aos EUA, bodas de vovó se incluem em eventos pessoais. A união de eventos sociais e pessoais forma parte dos eventos colaborativos, visto que são criados pelo usuário da aplicação. Na figura 1.1, é retratado o diagrama que representa os três tipos de eventos.

1.2 Objetivo

Considerando-se o contexto de organização de coleções de fotografias e a importância da localização e dos eventos relacionados a elas, este trabalho tem como objetivo propor um ambiente que vise não só aumentar o número de anotações corretas e automáticas das técnicas de anotação de localização, mas também diminuir a dependência de metadados existente nas técnicas de detecção de eventos por meio da

proposição de técnicas de anotação de localização e eventos. Além disso, este trabalho centra esforços em propor técnicas que dispensem elevado poder de processamento, razão pela qual tais técnicas não usarão informações de baixo nível.

1.2.1 Objetivos específicos

Os objetivos específicos são a proposição de técnicas de anotação de localização; a proposição de uma técnica de detecção de eventos; a validação das técnicas propostas; a implementação de um protótipo, denominada de PG++, que valida o ambiente proposto. A seguir, os objetivos específicos são detalhados.

1.2.2.1 Anotação de localização

A maioria das ferramentas de gerenciamento de coleções de fotografias auxilia a anotação da localização da fotografia, porém não o faz automaticamente. Dessa forma, uma técnica de anotação automática de localização será proposta e avaliada com a base de dados do PhotoGeo. O objetivo na proposição dessa nova técnica é encontrar a localização de alguma fotografia que ainda não a possua, com maior cobertura.

1.2.2.2 Detecção de eventos

Considerando-se o contexto de coleções de fotografias pessoais, existem técnicas de detecção de eventos sociais englobando fotografias de uma mesma câmera, de câmeras diferentes e também de redes sociais. Tais técnicas fazem uso de muitos metadados associados, como o relacionamento entre as pessoas, as pessoas anotadas, o contexto da fotografia. Dessa forma, visa-se propor uma técnica de detecção de eventos sociais em coleções de fotografias que seja menos atrelada aos metadados associados a elas.

1.2.2.3 Validação das técnicas propostas

A validação das técnicas é algo fundamental, visto que, para garantir a representatividade dos resultados, são necessários testes estatísticos. Assim sendo, as técnicas propostas para propagação de localização em fotografias pessoais serão testadas com a base de dados do PhotoGeo, que tem aproximadamente 7.900 fotografias georreferenciadas e com as pessoas presentes anotadas.

A base de dados usada para validar a detecção de eventos foi a proposta por Reuter et al. (2014). É composta por fotografias com eventos anotados, oriundas do

Flickr¹. Apenas as fotografias dos eventos que possuem pelo menos uma fotografia com localização foram consideradas. Dessa forma, a base filtrada resulta em 6.650 eventos e 131.551 fotografias. As fotografias presentes nessa base filtrada possuem *tags* textuais, data e hora da captura e localização geográfica na forma de latitude e longitude.

1.3 Contribuições

Neste trabalho, as principais contribuições são:

1. A proposição de duas técnicas de propagação de localização;
2. A realização de seleção automática de técnicas de propagação de localização com base no perfil do usuário e/ou da base de dados;
3. A proposição de uma técnica de detecção de eventos sociais;
4. A construção do PG++: um protótipo de ferramenta *web* para gerenciamento de fotografias;
 - 4.1. Agrupamento de fotografias para minimizar a quantidade que necessita de anotação manual;
 - 4.2. Organização automática em álbuns de fotografias geográficos.

1.3.1 Publicações

Este trabalho resultou nas seguintes publicações:

1. New Approaches for Geographic Location Propagation in Digital Photograph Collections
ICEIS 2014 - 16th International Conference on Enterprise Information Systems
Local: Lisboa, Portugal
2. Towards Better Propagation of Geographic Location in Digital Photo Collections
IBERAMIA 2014 - 14th Ibero-American Conference on Artificial Intelligence
Local: Santiago, Chile
3. PG++: Uma ferramenta para a anotação da localização em fotografias digitais
Best paper award no Workshop de Ferramentas e Aplicações (WFA) -

¹ www.flickr.com

WebMedia 2014

Local: João Pessoa, Paraíba

Além das publicações citadas, esta pesquisa resultou os seguintes trabalhos ainda em desenvolvimento:

1. Um *survey* em anotação de fotografia;
2. Artigo com a técnica de detecção de eventos sociais.

1.4 Estrutura da dissertação

O restante desta dissertação está organizado da seguinte forma: no capítulo 2, é apresentada a fundamentação teórica, que aborda a anotação de localização e eventos em fotografias, algoritmos de classificação, agrupamento e indexação. No capítulo 3, são mostrados os trabalhos relacionados a esta pesquisa. No capítulo 4, as técnicas de propagação de localização e a seleção automática são apresentadas, e a validação das técnicas e da seleção automática é descrita. No capítulo 5, a técnica de detecção de eventos sociais é explicada e a validação é descrita. No capítulo 6, a ferramenta *web PG++* é detalhada. Finalmente, no capítulo 7, estão descritas as conclusões e os trabalhos futuros.

Capítulo 2 – Fundamentação teórica

O objetivo deste capítulo é abordar os principais conceitos utilizados nesta pesquisa. Para isso, ele está dividido em quatro seções, que discorrem, respectivamente, sobre anotação em fotografia, algoritmos de classificação, agrupamento e indexação de texto. Ao fim, algumas considerações são apresentadas.

2.1. Anotação de fotografias

Nos tempos hodiernos, boa parte da organização de fotografias se processa de maneira automática por meio de sistemas *web* e *softwares* específicos para esse gerenciamento, presentes em máquinas pessoais. A organização é feita em álbuns e pode englobar os aspectos de tempo e lugar relacionados com as fotografias. Um álbum, por exemplo, pode conter fotografias que correspondem a um período de tempo ("verão de 2015"), um evento ("RockInRio") ou um lugar ("Europa") (MATELLANES et al., 2006). Os organizadores de fotografias fazem uso de informações atreladas a elas, os metadados. E, para tornar essa organização automática eficiente, é preciso manter esses metadados sempre atualizados. Com o aumento do número de fotografias, manter essa atualização não é fácil. O uso de abordagens para realizar a anotação dos metadados automaticamente pode facilitar a organização e recuperação de fotografias.

A anotação em fotografia consiste na associação de metadados com as fotografias armazenadas e enfatiza os aspectos significativos das imagens para a recuperação de memórias de eventos, festas, pessoas, momentos marcantes. Ademais, o uso de anotação permite que ferramentas de gerenciamento de fotografias sejam aprimoradas e diminuam o trabalho do usuário final. Tome-se como exemplo o fato de que a anotação da localização geográfica a que a fotografia pertence permite uma organização direta em relação ao local e uma apresentação da fotografia em uma interface de mapa. Essa abordagem possibilita a visualização de fotografias capturadas próximas a um determinado local, dado o nome ou as coordenadas do lugar. Além da localização, a anotação de eventos pode facilitar a busca de fotografias por épocas

características como "carnaval" ou "páscoa". Apesar da grande vantagem que a anotação pode trazer às ferramentas de gerenciamento de fotografias, o processo manual de anotação é custoso e demorado.

Em trabalhos na literatura, Naaman et al. (2004) e Sarvas et al. (2004) afirmam que a informação contextual para anotação automática é útil para organização e recuperação da fotografia, considerando-se, por exemplo, que saber a localização e data da captura possibilita descrevê-la bem antes da apresentação da imagem em si.

2.1.1 Contexto

A anotação automática pode ser feita com base no conteúdo da imagem e/ou no contexto em que a imagem está inserida. No escopo deste trabalho, o conteúdo da imagem são as informações de cada *pixel* da fotografia, enquanto o contexto é definido como algo que descreve a fotografia em si. A definição de contexto adotada é a de Dey e Abowd: "*Contexto é qualquer informação que pode ser utilizada para caracterizar a situação de uma entidade. Uma entidade é uma pessoa, lugar ou objeto que é considerado relevante para a interação entre o usuário e a aplicação, incluindo o próprio usuário e aplicação*" (DEY; ABOWD, 2000). Dessa forma, todos os metadados relacionados às fotografias são considerados contexto, tais como localização, evento, pessoas.

Uma fonte importante de informação de contexto das imagens é o arquivo EXIF (*EXchangeable Image File*) (CP-3451, 2002). É uma forma de armazenar metadados bastante comum e está presente na maioria das câmeras fotográficas do mercado. Esse tipo de armazenamento descreve informações, como nome, tamanho, data/hora, modelo da câmera, coordenadas geográficas. Segundo Viana et al. (2007), o grupo *W3C Multimedia Annotation* apontou a falta de sintaxe/semântica e a anotação automática de conteúdo como um dos principais problemas no desenvolvimento de ferramentas de gerenciamento de fotografias. O contexto descrito de maneira formal e explícita permite a inferência de novas informações de contexto. Viana et al. (2007) também afirma que os desafios para ferramentas que lidam com contexto são planejar e desenvolver sistemas que possam adaptar-se aos usuários e a suas preferências. Ante o exposto, este trabalho centra esforços nas informações contextuais de localização geográfica e dos eventos relacionados às fotografias.

2.1.2 Localização

Neste trabalho, a anotação automática de localização se dá por meio da propagação de localização. Segundo o dicionário Aurélio (FERREIRA, 2004), propagação é a ação ou efeito de propagar, que significa transportar-se através de algo. Assim sendo, este trabalho se preocupa em transportar a localização de uma fotografia para outra através da similaridade entre elas.

A localização de uma fotografia pode ser armazenada como coordenadas geográficas (latitude e longitude), nome do lugar ("João Pessoa"), descrição pessoal do lugar ("Minha casa") ou até com base em alguma referência ("em frente à Torre Eiffel"). Além disso, a localização pode, ainda, representar a posição da câmera ou a posição do objeto fotografado. Uma fotografia da estátua da Liberdade, por exemplo, pode ser feita de longe e a posição da câmera pode não corresponder à posição da estátua.

Neste trabalho, a localização de uma fotografia é entendida como a posição da câmera fotográfica no momento da captura. Essa posição pode ser captada pela própria câmera como uma coordenada geográfica. A posição é tratada como uma coordenada geográfica, pois, tendo como referência a localização, é possível obter o nome do lugar e associar a coordenada a descrições pessoais de usuários. Essa associação da coordenada com o nome do lugar e/ou descrições pessoais de usuários pode ser coletada por serviços externos, como o Wikimapia.

A captura da coordenada geográfica realizada pelas câmeras fotográficas atuais é feita através do GPS, logo, por limitações do sistema, fotografias capturadas no mesmo local podem receber coordenadas diferentes. A anotação automática através de propagação necessita de outras fotografias com localização para que alguma localização seja propagada para as fotografias que não possuem localização.

Como pode existir uma variação na localização inserida pelo GPS nas fotografias, para identificar se a anotação de localização foi correta ou não, é necessário considerar um limiar espacial. Uma anotação será considerada correta quando a coordenada estiver a uma distância espacial menor ou igual ao limiar considerado. Destarte, este trabalho adota o limiar espacial de 100 metros, pois foi o menor encontrado na literatura (LACERDA et al., 2013). Esse limiar foi usado,

capturarem as fotografias, as pessoas não costumam realizar anotações. Além disso, as ferramentas não auxiliam essa anotação direta dos eventos.

Neste trabalho, a anotação automática de eventos é realizada com base nas informações contextuais sobre a localização e o tempo das fotografias. Com a facilidade em encontrar dispositivos com câmeras fotográficas, tornou-se comum, em eventos sociais, a presença de inúmeras câmeras fotográficas de pessoas distintas. A captura de várias fotografias é uma situação comum em eventos sociais, o que torna a propagação de localização viável. Por isso, no escopo da anotação de eventos sociais nesta pesquisa, é assumido que, antes da anotação automática de evento social, as fotografias recebem a anotação de localização e trazem as coordenadas geográficas associadas.

2.1.4 Métricas

Neste trabalho, foram consideradas quatro métricas encontradas na literatura: precisão, cobertura, acurácia e medida F (RIJSBERGEN, 1979). A precisão indica a fração de anotações corretas em relação a todas as anotações realizadas. A cobertura representa a fração de anotações corretas em relação a todas as anotações que deveriam ter sido feitas. A acurácia aponta o número de anotações que deveriam ser feitas menos o número de anotações erradas em relação ao número de anotações que deveriam ser feitas. A medida F normaliza a combinação entre precisão e cobertura em um intervalo de 0 a 1, em que 0 é o pior resultado e 1 o melhor resultado.

As métricas adotadas são definidas com base em três números: “Sc”, que representa a quantidade de anotações corretas, “Se”, que indica a quantidade de anotações erradas, e “Sn”, indicando a quantidade de fotografias que não receberam propagações. As métricas são calculadas da seguinte forma:

- **Cobertura:**
$$\frac{Sc}{(Sc+Se+Sn)}$$
 (2.1)

- **Precisão:**
$$\frac{Sc}{(Sc+Se)}$$
 (2.2)

- **Acurácia:**
$$\frac{(Sc+Sn)}{(Sc+Se+Sn)}$$
 (2.3)

- **Medida F:**
$$2 \cdot \frac{Precisão \cdot Cobertura}{Precisão + Cobertura}$$
 (2.4)

2.2. Algoritmos

Nesta seção, são explicados cinco algoritmos utilizados nesta pesquisa: regressão linear múltipla (FREEDMAN, 2009), fatoração de matrizes (MEYER, 2000), máquina de vetores de suporte (MITCHELL, 1997), agrupamento DBScan (ESTER et al., 1996) e indexação TF-IDF (RAJARAMAN; ULLMAN, 2011).

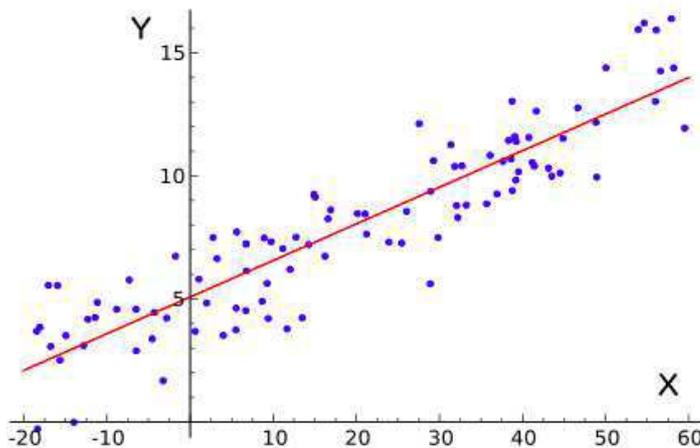
2.2.1 Regressão linear múltipla

Para explicar o conceito de regressão linear múltipla, é necessário entender o conceito de regressão linear simples. No aprendizado de máquina, regressão linear simples é uma abordagem para modelar o relacionamento entre uma variável e outra. Nela, os dados são modelados para se estimar o valor esperado de uma variável Y, dados os valores de outra variável X. Em geral (e no contexto deste trabalho), trata-se da questão de se estimar um valor condicional esperado.

A regressão linear simples é denominada "linear" em virtude de o relacionamento das variáveis Y e X ser modelado em uma função linear. A relação entre Y e X pode ser formalizada por meio da equação 2.5, em que α é uma constante que representa a interceptação da reta com o eixo vertical, β é o coeficiente angular da reta e ε é um valor de correção residual e possíveis erros de medição.

$$Y = \alpha + \beta.X + \varepsilon \quad (2.5)$$

Figura 2.2. Regressão linear simples



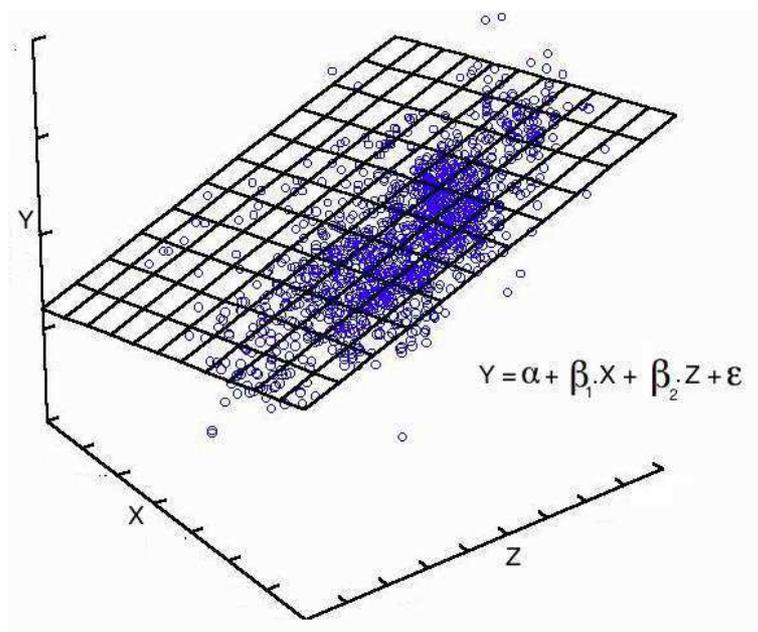
Na figura 2.2, é apresentado um exemplo de regressão linear. Cada ponto azul representa um par (X,Y) da base de dados usada para criar a regressão linear. O eixo vertical representa a variável Y e o eixo horizontal, a variável X. Dessa forma, a linha

vermelha é a função linear que representa a relação das variáveis Y e X. Essa função linear é o resultado da regressão. Com a função linear, é possível estimar o valor de Y em função de algum valor de X.

A regressão linear múltipla é semelhante à regressão linear simples, porém a relação ocorre de uma variável para múltiplas variáveis, como o nome sugere. Assim, a regressão relacionará uma variável Y com um conjunto de variáveis através de uma função linear. Essa relação com múltiplas variáveis pode ser formalizada por meio da equação 2.6:

$$Y = \alpha + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \dots + \beta_n \cdot X_n + \varepsilon \quad (2.6)$$

Figura 2.3. Regressão linear múltipla



Na figura 2.3, é apresentado um exemplo de regressão linear múltipla, em que a variável Y é relacionada com as variáveis X e Z. Cada ponto azul representa uma tupla (X,Z,Y). O eixo vertical representa a variável Y, o eixo horizontal, a variável Z e o eixo de profundidade, a variável X. Dessa forma, o plano quadriculado é o resultado da regressão linear múltipla e representa a função linear de relação de Y com X e Z. Com a função linear de resultado, é possível estimar o valor de um Y qualquer com base nos valores de X e Z.

2.2.2 Fatoração de matriz

Uma matriz bidimensional é uma tabela de m linhas e n colunas. A fatoração de matriz, como o nome sugere, divide a matriz em um produto de matrizes. Seja A uma matriz m por n , a fatoração de matriz encontrará as matrizes X e Y , tal que:

$$A_{m \times n} = X_{m \times r} \cdot Y_{r \times n} \quad (2.7)$$

Como apresentado na equação 2.7, a matriz A foi fatorada em um produto de X e Y . Os algoritmos de fatoração de matrizes existentes encontram o valor de r automaticamente e gera as matrizes X e Y como resultado. Dependendo da matriz, a fatoração pode gerar um valor aproximado e não o real.

A fatoração de matrizes pode ser usada para estimar valores por meio da matriz Y . Seja A uma matriz em que as colunas $[1, \dots, n-1]$ são valores conhecidos e a coluna n é a variável a ser estudada. A matriz A representa o conjunto de treinamento, a matriz X representa os valores usados como parâmetro e a coluna $n-1$ de A representa os valores que precisam ser estimados pelo modelo.

A fatoração usada para estimar o valor de uma coluna da matriz é feita para encontrar a matriz Y , pois a matriz X é composta pelas demais colunas. Dessa forma, pode não existir uma matriz Y , em que a multiplicação com X resulte na matriz A . Então, é encontrada a matriz Y , em que a multiplicação por X resulte em uma matriz próxima de A .

Na figura 2.4, é apresentado um exemplo de fatoração de matrizes para a estimação de valores. Considere-se a matriz A' , em que se deseja estimar os valores da coluna 4 para as linhas verdes. Como os valores da coluna são desconhecidos, os campos são preenchidos com o valor 0. A matriz X' é composta pelas colunas 1, 2 e 3, e a fatoração encontra a matriz Y' , em que faz a multiplicação com X' gerar uma matriz próxima de A' .

Com a matriz Y' e as linhas verdes da matriz X' , é possível encontrar o valor estimado para os valores ausentes na matriz A' (células em vermelho com 0). Dessa forma, como a matriz Y' é calculada para gerar o menor erro, os valores estimados podem ser usados como previsão de um valor que ainda não foi preenchido.

Figura 2.4. Estimação de valores com fatoração de matriz

$$\begin{array}{c}
 \mathbf{A} \\
 \left[\begin{array}{cccc}
 a_{11} & a_{12} & a_{13} & a_{14} \\
 a_{21} & a_{22} & a_{23} & a_{24} \\
 a_{31} & a_{32} & a_{33} & a_{34} \\
 a_{41} & a_{42} & a_{43} & a_{44}
 \end{array} \right] \approx \begin{array}{c}
 \mathbf{X} \\
 \left[\begin{array}{ccc}
 a_{11} & a_{12} & a_{13} \\
 a_{21} & a_{22} & a_{23} \\
 a_{31} & a_{32} & a_{33} \\
 a_{41} & a_{42} & a_{43}
 \end{array} \right] \cdot \begin{array}{c}
 \mathbf{Y} \\
 \left[\begin{array}{cccc}
 y_{11} & y_{12} & y_{13} & y_{14} \\
 y_{21} & y_{22} & y_{23} & y_{24} \\
 y_{31} & y_{32} & y_{33} & y_{34}
 \end{array} \right]
 \end{array}
 \end{array}$$

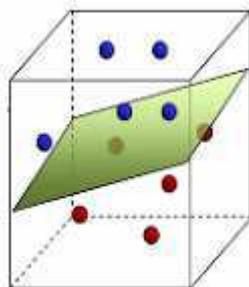
$$\begin{array}{c}
 \mathbf{A}' \\
 \left[\begin{array}{cccc}
 a_{11} & a_{12} & a_{13} & a_{14} \\
 a_{21} & a_{22} & a_{23} & a_{24} \\
 a_{31} & a_{32} & a_{33} & a_{34} \\
 a_{41} & a_{42} & a_{43} & a_{44} \\
 a_{51} & a_{52} & a_{53} & 0 \\
 a_{61} & a_{62} & a_{63} & 0 \\
 a_{71} & a_{72} & a_{73} & 0
 \end{array} \right] \approx \begin{array}{c}
 \mathbf{X}' \\
 \left[\begin{array}{ccc}
 a_{11} & a_{12} & a_{13} \\
 a_{21} & a_{22} & a_{23} \\
 a_{31} & a_{32} & a_{33} \\
 a_{41} & a_{42} & a_{43} \\
 a_{51} & a_{52} & a_{53} \\
 a_{61} & a_{62} & a_{63} \\
 a_{71} & a_{72} & a_{73}
 \end{array} \right] \cdot \begin{array}{c}
 \mathbf{Y}' \\
 \left[\begin{array}{ccccccc}
 y_{11} & y_{12} & y_{13} & y_{14} & y_{15} & y_{16} & y_{17} \\
 y_{21} & y_{22} & y_{23} & y_{24} & y_{25} & y_{26} & y_{27} \\
 y_{31} & y_{32} & y_{33} & y_{34} & y_{35} & y_{36} & y_{37}
 \end{array} \right]
 \end{array}
 \end{array}$$

2.2.3 Máquina de vetores de suporte (SVM)

Máquina de vetores de suporte (SVM) está inserida em uma área de inteligência artificial chamada aprendizado de máquina, cujo objetivo é a construção de sistemas capazes de adquirir conhecimento de forma automática. O aprendizado relacionado a um problema é realizado por meio da análise de experiências armazenadas que obtiveram uma solução correta para o problema em questão. Esse aprendizado gera conhecimento, que pode ser usado para resolver problemas semelhantes no futuro. Dentre as técnicas de aprendizado, destacam-se as de aprendizado supervisionado, as quais utilizam dados rotulados previamente. No aprendizado supervisionado, o modelo de classificação SVM é bastante utilizado em problemas de classificação.

SVM é um método de classificação que usa aprendizagem de máquina com o objetivo de encontrar um hiperplano de alta dimensão para separar classes. O hiperplano é calculado com base na minimização do risco estrutural envolvido. SVM procura um hiperplano representado por vetores, o qual divide os dados de treinamento nas classes desejadas (ex.: azul e vermelho). A ideia básica do algoritmo de treinamento é encontrar um hiperplano, representado pelo vetor que separa os dados armazenados em classes pelo espaço mais amplo possível.

Figura 2.5. Hiperplano encontrado com SVM



Na figura 2.5, é apresentado um exemplo de hiperplano encontrado por meio de SVM para classificar os objetos em azul ou vermelho. O hiperplano resultante é um quadrilátero plano, representado pela área verde, que conseguiu dividir completamente os dados em todo o universo analisado. O hiperplano criado para classificar os dados existentes é utilizado para classificar novos valores. Dessa forma, ao chegar uma nova instância no universo do problema, o hiperplano é usado para classificar e indicar a classe adequada para aquela instância.

2.2.4 Indexação TF-IDF

Um documento de texto contém vários termos, frases e palavras-chave que podem ser empregados para recuperá-lo. Da mesma forma, é preciso usar algum elemento de indexação para melhorar a busca por documentos existentes em uma base de dados. Para tal, a indexação TF-IDF foi criada. Essa indexação considera a frequência do termo (*term frequency* - *TF*) e a frequência de documentos inversa (*inverse document frequency* - *IDF*).

A frequência dos termos nos documentos pode ser calculada de maneira bruta ou normalizada. Na indexação TF-IDF usada neste trabalho, é considerado o cálculo normalizado. Seja $qtd(t,d)$ a função que retorna a quantidade de vezes que o termo t qualquer aparece em um documento d qualquer. A função $TF(t,d)$, que encontra o valor TF de um termo t em um documento d , é definida como:

$$TF(t,d) = \frac{qtd(t,d)}{|d|} \quad (2.8)$$

A frequência de documentos inversa é calculada para os termos isolados com base no número total de documentos e no número de documentos que contém um determinado termo. Seja D o conjunto de documentos e $numD(D,t)$ a função que

encontra o número de documentos em D , que contém o termo t . A função $IDF(t,D)$, que encontra o valor IDF de um termo t em um conjunto de documentos D , é definida como:

$$IDF(t,D) = 1 + \log_e\left(\frac{|D|}{numD(D,t)}\right) \quad (2.9)$$

Dessa forma, a função $TF.IDF(t,d)$ de um termo t em relação a um documento d pertencente a D é definida pela equação 2.10:

$$TF.IDF(t,d) = TF(t,d) \times IDF(t,D) \quad (2.10)$$

2.2.5 Similaridade de cosseno com TF-IDF

Para facilitar a recuperação dos documentos, é feito o cálculo de similaridade de cosseno entre documentos, usando a função $TF.IDF$. Seja D o conjunto de todos os documentos, d_i um documento qualquer, d_{ij} o termo j do documento d_i e $d_i[t]$ o $tf-idf$ do termo t em relação ao documento d_i . A similaridade de cosseno é calculada entre documentos por meio das funções $dxd(d_1,d_2)$ e $raiz2(d_1)$, definidas pelas equações 2.11 e 2.12, respectivamente.

$$dxd(d_1,d_2) = \sum_{k=1} (d_1[d_1t_k] \times d_2[d_1t_k]) \quad (2.11)$$

$$raiz2(d_1) = \sqrt{\sum_{k=1} (d_1[d_1t_k])^2} \quad (2.12)$$

Com essas funções, a similaridade de cosseno $simCoss(d_1,d_2)$ é definida pela equação 2.13.

$$simCoss(d_1,d_2) = \frac{dxd(d_1,d_2)}{raiz2(d_1) \times raiz2(d_2)} \quad (2.13)$$

A recuperação de documentos com base em busca textual considera o texto inserido como um documento e, então, calcula o $tf-idf$ dos termos da busca. Em seguida, calcula a similaridade da busca com todos os documentos. O cálculo é feito por meio da função $simCoss(d_1,d_2)$, em que d_1 representa a busca textual e d_2 representa um documento qualquer pertencente à base de dados. As equações 2.8, 2.9, 2.10, 2.11, 2.12 e 2.13 são definidas por RAJARAMAN e ULLMAN (2011).

2.2.6 Agrupamento DBSCAN

O DBSCAN é um algoritmo de agrupamento baseado na densidade de *clusters*. Esse é um algoritmo de aprendizado não supervisionado com base em observação. O algoritmo separa a base de dados em grupos, de tal forma que cada grupo seja um conjunto de objetos, no qual cada objeto seja mais similar a outros dentro do mesmo grupo do que a qualquer outro fora do grupo.

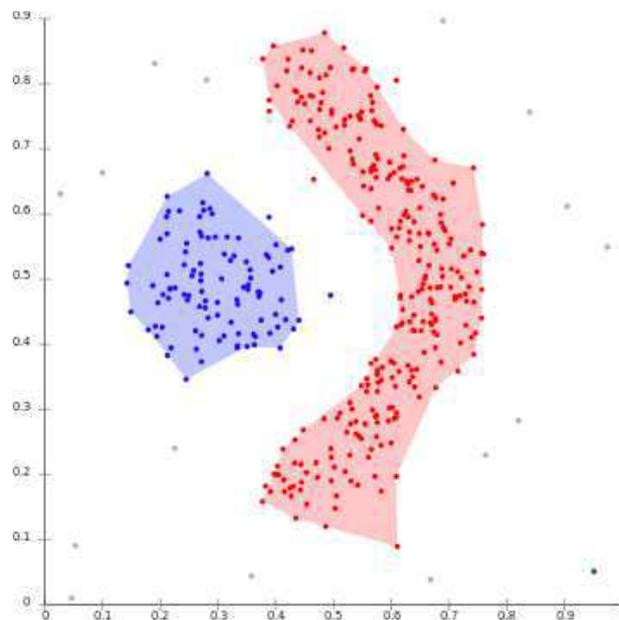
Esse algoritmo é usado para agrupar dados de maneira arbitrária e é robusto em relação a ruídos e *outliers*. Para que o DBSCAN funcione, é preciso definir dois parâmetros: *eps* e *minPts*. O parâmetro *eps* indica a distância máxima entre dois pontos pertencentes a um mesmo grupo e o parâmetro *minPts* representa a quantidade mínima de pontos que cada grupo deve conter. Dessa forma, o algoritmo encontrará os grupos adequados para a configuração de *eps* e *minPts* desejada.

Seja G o conjunto de grupos criado pelo DBSCAN, g_i um grupo qualquer de G e $dist(a,b)$ a função que calcula a distância entre os objetos a e b :

$$\forall g_i \in G \rightarrow |g_i| \geq minPts \quad (2.11)$$

$$\forall o_i, o_j \in g_i \rightarrow dist(o_i, o_j) \leq eps \quad (2.12)$$

Figura 2.6. Exemplo de agrupamento com DBSCAN



Na figura 2.6, é apresentado um exemplo de agrupamento com o algoritmo DBScan. Os polígonos coloridos representam os grupos criados. É claro que a anatomia dos grupos é arbitrária e construída com o acoplamento de novos elementos. Caso o algoritmo seja executado novamente, para a mesma base de dados e com as mesmas configurações de *eps* e *minPts*, os grupos serão sempre os mesmos. Caso a configuração de *eps* e *minPts* mude, ainda que pouco, o resultado pode ser diferente.

2.3. Considerações finais

Neste capítulo, apresentou-se a fundamentação teórica deste trabalho. Foram expostos os principais conceitos relacionados com anotação de localização e eventos em fotografias, aspectos relacionados a algoritmos de aprendizado de máquina, o funcionamento do algoritmo de agrupamento DBScan e da similaridade de cossenos entre documentos por meio de indexação TF-IDF.

Os algoritmos estudados e utilizados foram elencados com base na literatura. Foram escolhidos os algoritmos mais frequentes nos trabalhos analisados. No próximo capítulo, serão apresentados alguns trabalhos relacionados à pesquisa realizada.

Capítulo 3 – Trabalhos relacionados

Existem trabalhos na literatura que buscam melhorar a anotação em fotografias digitais, envolvendo a localização da fotografia e os eventos associados a ela. Neste capítulo, é apresentado um levantamento bibliográfico das pesquisas sobre anotação de localização geográfica e anotação de eventos.

3.1 Localização geográfica

Gong et al. (2011) trabalham com a ideia de prever a localização de pessoas para melhoria de serviços de telefonia móvel. A ideia proposta usa um modelo de correlação social que toma como base as localizações dos contatos das pessoas. O modelo é simples e pode ser facilmente adaptado ao contexto de fotografias, no qual a correlação se torna a presença em uma mesma fotografia. Contudo, é dependente de anotações de pessoas e não considera uma diferença temporal grande (dias ou semanas) entre uma anotação e outra.

Gao et al. (2012) abordam o assunto da previsão da localização como um todo para solução de problemas, como planejamento de tráfego, propaganda direcionada, dentre outros. É proposto um modelo para prever a localização do usuário, baseando-se em um histórico de visitas e levando-se em conta o contexto espaço-temporal. Trata-se de um modelo que também pode ser adaptado ao contexto de fotografias, no qual o histórico de visitas é construído pelas fotografias com localização. Entretanto, uma simples mudança na rotina do usuário (uma viagem a um lugar nunca visitado, por exemplo) fará com que a técnica não encontre a localização.

Hays e Efros (2008) propuseram um algoritmo chamado "im2gps", que estima a localização geográfica de uma fotografia com base na localização geográfica de fotografias com alta similaridade visual. Como o "im2gps" usa uma base de fotografias, a localização encontrada é estritamente ligada à quantidade de fotografias na base de dados e depende de processamento de imagem que consome bastante tempo. Além disso, a técnica considera um raio de distância da localização real de

quilômetros, que, para o contexto de fotografias pessoais, é muito grande.

Ivanov et al. (2012) propuseram a propagação de *geotags* baseada na combinação de detecção de objetos repetidos. A ideia é a propagação de *geotags* usando outras fotografias com *geotags*. Essa abordagem pode se tornar frágil quando usada para coleções de fotografias pessoais, pois, em vários momentos, as pessoas presentes na fotografia cobrem a maior área da imagem. Considerando-se que pessoas não representam uma localização, as ideias de similaridade visual e detecção de objetos repetidos podem gerar resultados que não correspondam à realidade.

Hollenstein e Purves (2013) sugerem que a localização geográfica deve ser definida por meio da maneira como os usuários descrevem o lugar, em vez de latitude e longitude. Assim, *tags* como "Cristo Redentor" deveriam ser adicionadas à localização. Relativamente à implantação de *tags* associadas à localização geográfica, é temeroso garantir consistência, quando as *tags* são definidas pelo usuário, e sugerir, quando não são definidas pelo usuário.

Lacerda et al. (2013) abordam inconsistências nas localizações de fotografias e também na propagação de localização. A ideia proposta usa um agrupamento temporal para encontrar a localização por meio da fotografia mais próxima no tempo e apresenta bons resultados. Entretanto, uma coleção de fotografias com grande intervalo de tempo entre elas torna impreciso o agrupamento temporal, resultando em localizações distantes da realidade.

Neste trabalho, são propostas duas técnicas de propagação de localização e uma seleção automática de técnicas de propagação. As técnicas de propagação são direcionadas a dois tipos de usuários que os trabalhos relacionados não abrangem: usuários que viajam muito e usuários que frequentam inúmeros eventos. Além disso, a técnica de seleção automática permite que outras técnicas de propagação sejam usadas para atender outros comportamentos de usuários.

3.2 Eventos

CrEve (ZIGKOLIS et al., 2012) é um *framework* de anotação colaborativa de eventos. O trabalho une vários tipos de indexação, permitindo a busca por texto e

contexto da imagem. O *framework* gerou melhora na qualidade das anotações dos eventos e também na visualização das imagens. Faz uso de processamento de imagens para identificar as fotografias com similaridade visual. A base de dados foi coletada pelos autores, apesar de citar bases de dados existentes. Essa base é composta por fotografias das redondezas de Paris e Barcelona, que possuem os seguintes metadados: localização, tempo em que a imagem foi capturada, usuário e identificador. Contudo, não faz a detecção de novos eventos, ou seja, a detecção se processa apenas para eventos cadastrados pelo usuário.

Lim et al. (2003) apresentam um modelo de detecção de eventos baseado em processamento de imagem que compara as imagens e, de acordo com a similaridade, relaciona-as ao respectivo evento. Tal modelo apresenta bons resultados, porém os eventos são detectados como, por exemplo, piscina, parque. Esses eventos abstratos são previamente definidos pelo usuário, logo um novo evento nunca é criado.

Figueirêdo et al. (2012) apresentam uma técnica de detecção de eventos pessoais considerando fotografias de câmeras diferentes. A técnica faz uso das pessoas presentes na fotografia, da data e hora em que foi tirada e da localização geográfica. O trabalho se vale de metadados, e a técnica de detecção de eventos dispensa processamento de imagem, porém o grande número de metadados necessários para a obtenção de bons resultados torna a técnica dependente daqueles. Assim sendo, essa dependência pode gerar trabalho adicional para o usuário.

Brenner e Izquierdo (2012) apresentam um *framework* para detectar eventos sociais e recuperar fotografias em coleções colaborativas de fotografias digitais. Esse trabalho trata o problema da relação entre fotografias e eventos de maneira completa, pois faz desde a detecção até a recuperação de fotografias envolvidas nos eventos. O *framework* prescinde de processamento digital de imagens e trabalha apenas com os metadados: localização, informação textual associada, data e horário das fotografias. A informação textual associada é tratada com processamento de linguagem natural para extrair palavras-chave e possíveis tópicos que possam identificar o evento da fotografia. Usa fontes externas como *Google Geocode*, *Geonames*, *DBpedia* e *Last.fm* para auxiliar a expansão de conteúdo associado e, conseqüentemente, o funcionamento do *framework*. O trabalho dos autores apresentou bons resultados, porém os eventos obtidos não correspondem aos eventos reais. A detecção ocorre no nível de categoria

(esporte, música, festa). Um evento como "Copa do mundo de futebol" seria enquadrado na categoria esporte.

Rabbath et al. (2012) concentram esforços em aplicações de fotografias baseadas na rede social *Facebook*. Lidam com os metadados fornecidos pelo *Facebook* e apresentam uma solução para detectar fotografias que pertencem ao mesmo evento e que estão distribuídas entre vários amigos. Rabbath et al. usam um modelo baseado em processamento digital de imagens, *tags*, relacionamentos da rede social e estruturas da própria rede social. A solução, entretanto, é dependente da rede social, não considera a localização das fotografias e não engloba eventos sociais diretamente.

Nakaji e Yanai (2012) propõem um modelo para agrupar e selecionar fotografias relevantes de um mesmo evento detectadas pela rede social *Twitter*. A solução coleta fotografias georreferenciadas, faz uma extração de características visuais e, então, aplica o modelo de agrupamento e seleção de fotografias representativas. O modelo proposto obteve êxito, porém é necessário que o evento desejado seja previamente fornecido. Assim sendo, a solução não detecta automaticamente os eventos.

Feng et al. (2014) abordam a presença de eventos em redes sociais de maneira diferente. Em vez de procurar os eventos nas redes sociais, o trabalho desses autores busca pessoas influentes que podem ser organizadores de eventos. Foram propostas duas técnicas gulosas e uma técnica aprimorada, que garante encontrar a melhor resposta, caso exista. Todas as soluções propostas por Feng et al. fazem uso do grafo que representa a rede social. Apesar de o trabalho não englobar a detecção dos eventos, apresenta mais de uma solução para encontrar as pessoas influentes. Dessa forma, por intermédio dessas pessoas, os eventos podem ser encontrados.

Ilina et al. (2012) trabalham com detecção de eventos na rede social *Twitter*. O trabalho dos autores consiste em identificar se uma determinada publicação está ou não relacionada com um evento. Dessa forma, o processo de detecção de eventos torna-se binário, logo a solução proposta não identifica possíveis eventos específicos. Ela divide o texto em *n-grams* e calcula a probabilidade de cada *n-gram* estar relacionado com um evento. Com todas as probabilidades em mãos, calcula-se, por meio do teorema de Bayes, a probabilidade de o texto completo estar relacionado com um evento. A solução proposta no trabalho apresenta bons resultados e indica ser bastante

promissora, contudo não identifica os eventos aos quais as publicações estão relacionadas. Cumpre registrar que, mesmo não identificando o evento, a solução pode ser usada como filtro para evitar processamento de texto desnecessário.

Vyas et al. (2013) realizaram experimentos para identificar algumas práticas envolvendo eventos em fotografias. Foram feitas três abordagens para estudar eventos em fotografias: entrevistas, visitas guiadas e observação de eventos reais. As entrevistas foram feitas por meio de um questionário *online*, no qual os participantes podiam responder a algumas perguntas sobre seus costumes envolvendo fotografias. Nas visitas guiadas, não só foram feitas perguntas similares às do questionário *online*, mas também foi analisada a maneira como as famílias apresentavam as fotografias (computador, álbum impresso). A observação de eventos reais foi feita em abordagens a pessoas aleatórias em zoológicos, parques de diversão, parques públicos e eventos culturais, dirigindo-lhes as mesmas perguntas do questionário *online*. A observação sobre a representação de eventos mostrou que os de maior significado na vida dos entrevistados apresentam uma qualidade melhor no armazenamento. Além disso, algumas práticas foram identificadas para facilitar a representação dos eventos. A primeira prática é capturar as pessoas importantes daquele evento, que geralmente é representado por uma foto em grupo. A segunda é capturar os momentos significativos do evento, como soprar a vela de aniversário, dançar pela primeira vez. Outra prática comum é capturar aspectos que caracterizam o evento, como um momento difícil de uma viagem. Apesar da análise feita pelos autores, eles não apresentam nenhuma ferramenta focada nas práticas identificadas.

Mezaris et al. (2014) apresentam uma discussão sobre eventos da vida real em arquivos multimídia. Destacam e resumem nove trabalhos envolvendo saúde de idosos, fenômenos naturais (como tufões e ciclones) e detecção de eventos em vídeos e coleções de fotografias. O trabalho dos autores mostra a presença da detecção de eventos em diversas áreas e destaca a importância do contexto em fotografias e vídeos. Os trabalhos resumidos envolvendo fotografias estão focados em organização e recuperação de contexto. Contudo, os autores não propõem nenhuma abordagem para a detecção de eventos sociais.

Paniagua et al. (2013) propõem novas abordagens para detecção de eventos sociais. Essa detecção utiliza as informações de tempo e espaço das fotografias

presentes nos eventos pessoais dos usuários e se dá com funções de similaridade temporal e espacial. As funções trabalham com conjuntos de fotografias, os quais representam os eventos pessoais dos usuários. A função de similaridade temporal calcula o quão semelhante os eventos pessoais de usuários distintos são semelhantes com relação à data e horário em que as fotografias foram capturadas e a função de similaridade espacial com relação à localização das fotografias. O trabalho apresenta bons resultados, já que a detecção de eventos sociais alcança uma precisão de 78.76%; porém, para que um evento social seja detectado, deve existir interseção entre os eventos pessoais de usuários.

GESODECK (GAO et al., 2013) é uma ferramenta para detectar e acompanhar eventos sociais em redes sociais como o *Twitter*. A ferramenta detecta o comportamento do usuário na rede social e usa as informações geográficas do *post* para identificar a alternância de usuários que publicam conteúdo de uma área geográfica específica. Os autores definem que essa alternância de usuários ultrapassa o natural dessas regiões quando, em um determinado período de tempo, o número de usuários alternando publicações na região está maior que o padrão da região. Quando uma alternância de usuários é detectada além da média da região, aquele conjunto de publicações segue para processamento de linguagem natural. Após os eventos terem sido detectados, o acompanhamento ocorre com um agrupamento *K-means* (HARTIGAN, 1975), considerando-se contagens de republicações e palavras-chave. Entretanto, existem locais públicos, como a Torre Eiffel, em que a alternância de usuários é sempre bastante elevada, fazendo com que não ultrapasse a média padrão durante um evento no local.

Grabovitch-Zuyev et al. (2014) usam a localização e o texto presente em publicações de *microblogs* para encontrar correlações envolvendo essas informações. Usam dois principais parâmetros para correlacionar os usuários observados. O primeiro considera a distância geográfica, por meio de uma função que calcula a distância entre dois pontos (latitude e longitude), levando-se em conta a curvatura da terra e os conjuntos de localizações que representam cada usuário. A correlação do primeiro parâmetro é encontrada pela diferença média entre os pontos dos conjuntos que representam cada usuário. O segundo parâmetro considera uma similaridade textual baseada em repetição de palavras-chave, encontrada pela fórmula do *tf-idf*

(RAJARAMAN e ULLMAN, 2011). O trabalho mostra que existe correlação entre o conteúdo do texto e a área em que o usuário se encontra. Apesar de o trabalho comprovar a relação entre o texto e lugar, os autores não apresentam nenhuma abordagem para detectar eventos sociais.

Dao et al. (2013) propõem novas técnicas para detecção de eventos em coleções de fotografias digitais. Enfatizam a organização de fotografias em eventos, citam trabalhos focados em agrupar fotografias em eventos sociais e destacam a prática comum do usuário em associar suas fotografias com o que estava acontecendo no momento da captura. As técnicas apresentadas baseiam-se em combinações do uso de *keywords*, informação textual (título, descrição) e informações visuais. O banco de dados do *MediaEval* 2012 é usado e comparado com técnicas existentes usadas para solucionar o desafio de detecção de eventos sociais. O trabalho apresenta bons resultados, porém os eventos detectados são abstratos ("protestos em Madrid", "partidas de futebol na Espanha", dentre outros) e previamente determinados.

Psallidas et al. (2013) abordam a identificação de eventos não conhecidos em redes sociais, propondo uma nova técnica para essa identificação. Essa técnica utiliza informações associadas às publicações coletadas (usuário que publicou, data/hora, texto descritivo). Com essas informações, usa funções de similaridade temporal, espacial e textual para criar grupos de publicações. Além das funções de similaridade, os autores propõem o uso das *URLs* publicadas pelos usuários, de termos usados intensamente em um curto intervalo de tempo e da deterioração temporal para aperfeiçoar o agrupamento temporal. Os resultados encontrados pelos autores mostram que o uso de termos repetidos intensamente em um curto intervalo de tempo, em conjunto com a deterioração temporal, melhora a eficácia na identificação de eventos. O uso de *URLs* só é recomendado pelos autores para redes sociais que contêm a maioria das publicações associadas a um *link*. Todavia, o trabalho é dependente das redes sociais, fazendo com que a detecção se dê somente com eventos sobre os quais muitos usuários falam.

Choudhury et al. (2012) iniciam trabalhos na identificação de tipos de usuários na rede social *Twitter*. Os autores concentram esforços nessa tarefa visando melhorar o acompanhamento de eventos divulgados na rede social. Classificam os usuários com base em informações, como número de amigos, quantidade de *retweets*, dentre outras.

Oito diferentes classificadores foram testados, e o *k-NN* (ALTMAN, 1992) foi o que melhor classificou os usuários. Com os resultados obtidos, os autores concluíram que a participação dos tipos de usuários varia de acordo com o evento, porém o trabalho não faz detecção de eventos.

Becker et al. (2011) apresentam três novas formas de selecionar conteúdo relevante associado a um determinado evento previamente conhecido. Os autores fizeram um estudo comparativo entre as novas abordagens sugeridas e identificaram que a seleção nomeada como "*centroid*" foi a que demonstrou melhores resultados. Ela usa uma função de similaridade cosseno para comparar as mensagens com os principais termos do evento na representação *tf-idf* (RAJARAMAN e ULLMAN, 2011). O estudo comparativo foi feito considerando-se as seguintes métricas: qualidade, usabilidade e relevância. Qualidade reflete o quão bem interpretado aquele determinado texto é para um ser humano. Relevância aponta o quão bem o texto representa uma informação do evento. Usabilidade indica o potencial de uma mensagem de texto em passar informações para alguém que esteja interessado em um determinado evento. O trabalho dos autores identifica uma abordagem eficiente para centralizar mensagens sobre eventos, porém as abordagens indicadas não fazem a detecção de eventos sociais ainda não descobertos.

Becker et al. (2012) propõem estratégias para busca de conteúdo associado a eventos previamente conhecidos. Essa busca engloba diversos *sites* de compartilhamento de multimídia, como *Flickr*, *Twitter* e *Youtube*. Os autores concentram as estratégias nas métricas de precisão e cobertura. A busca por conteúdo se dá pela combinação de título e nome do local do evento. Várias consultas são feitas com variações do texto que representa o título e o nome do local. Essas variações são feitas com o uso ou não do texto exato, uso ou não de *stopwords*, uso apenas do título. O trabalho apresentou bons resultados e deu início a uma nova abordagem na busca por conteúdo relacionado a eventos. Contudo, o trabalho, além de não fazer buscas para eventos não planejados, não faz detecção de eventos.

Brenner e Izquierdo (2013) apresentam um *framework* baseado no desafio de detecção de eventos sociais (SED) do *MediaEval* 2013. Os autores fazem pré-processamento dos metadados para propagação de localização e usam agrupamentos temporais e espaciais para montar o conjunto de treinamento. O *framework* apresenta

bons resultados e faz desde a detecção até a recuperação de fotografias relacionadas a eventos, porém os eventos considerados são apenas "concerto de música" ou "partida de esporte". Dessa forma, a detecção não encontra o nome do evento ou as principais palavras-chave relacionadas a ele.

Neste trabalho, são propostas duas técnicas: uma de agrupamento de fotografias de um mesmo evento e outra de anotação automática de eventos em fotografias. A técnica de agrupamento usa as informações presentes nos metadados da fotografia para reunir fotografias de um mesmo evento, considerando-se a localização e o instante de tempo associados à fotografia. Por sua vez, a técnica de anotação automática permite que fotografias com ou sem localização recebam o evento associado. Além de a técnica de agrupamento ser independente de redes sociais externas, a técnica de anotação automática pode considerar eventos reais, definidos pelo usuário, e eventos desconhecidos.

3.3 Comparativo dos trabalhos relacionados

Nesta seção, é apresentada uma comparação entre os trabalhos acadêmicos estudados que tratam da anotação de localização ou eventos em coleções de fotografias digitais. No eixo de anotação de localização, os trabalhos propõem soluções que resolvem uma parcela específica do problema. Neste trabalho, as técnicas de anotação de localização são comparadas considerando-se as seguintes características: anotação automática de localização, uso de propagação de localização e anotação de localização autoadaptável ao perfil do usuário. Apesar de apresentarem bons resultados, os trabalhos na área de anotação de localização, além de não possuírem todas as características analisadas, não apontam um meio para lidar com elas.

Quanto ao eixo de anotação de eventos, os trabalhos estudados estão divididos nos seguintes grupos:

- Detecção de eventos abstratos - não detectam o evento real, o evento detectado é abstrato ("esporte", "concerto de música", dentre outros);
- Agrupamento de fotografias de um mesmo evento - a detecção de eventos ocorre apenas na forma de um agrupamento de fotografias, o evento real (nome, local) não é identificado;
- Eventos planejados - trabalha com eventos, porém não considera eventos ainda não cadastrados no sistema;
- Dependentes de redes sociais externas - o trabalho é construído para uma ou

mais redes sociais específicas fora do sistema proposto pelos autores.

É apresentado, na tabela 3.1, um estudo comparativo, analisando as seguintes características:

- C1.** Anotação automática de localização;
- C2.** Propagação de localização;
- C3.** Propagação de localização autoadaptável ao perfil da base de dados;
- C4.** Anotação de eventos;
- C5.** Eventos reais;
- C6.** Eventos definidos;
- C7.** Eventos desconhecidos;
- C8.** Independência de redes sociais externas;
- C9.** Agrupamento de conteúdo do mesmo evento.

Os trabalhos de Vyas et al. (2013) e Mezaris et al. (2014) não estão presentes na tabela comparativa por se tratarem de resumos de outros trabalhos e não proporem novas maneiras de resolver o problema de anotação de eventos. Com o estudo apresentado na tabela 3.1, fica evidente que as anotações de localização e de eventos não são abordadas de maneira completa. Além de nenhum trabalho apresentar uma análise das duas anotações juntas, há lacunas que precisam ser preenchidas.

Tabela 3.1. Estudo comparativo dos trabalhos estudados.

Trabalho	C1	C2	C3	C4	C5	C6	C7	C8	C9
GONG et al., 2011	✓	✓	✗	✗	✗	✗	✗	✓	✗
GAO et al., 2012	✓	✗	✗	✗	✗	✗	✗	✓	✗
HAYS e EFROS, 2008	✓	✓	✗	✗	✗	✗	✗	✓	✗
IVANOV et al., 2012	✓	✓	✗	✗	✗	✗	✗	✓	✗
HOLLENSTEIN e PURVES, 2013	✓	✗	✗	✗	✗	✗	✗	✓	✗
LACERDA et al., 2013	✓	✓	✗	✗	✗	✗	✗	✓	✗
ZIGKOLIS et al., 2012	✗	✗	✗	✓	✓	✓	✗	✓	✗
LIM et al., 2003	✗	✗	✗	✓	✗	✓	✗	✓	✓
FIGUEIRÊDO et al., 2012	✗	✗	✗	✓	✗	✓	✓	✓	✓
BRENNER e IZQUIERDO, 2012	✗	✗	✗	✓	✗	✓	✓	✓	✓
RABBATH et al., 2012	✗	✗	✗	✓	✗	✗	✗	✗	✓
NAKAJI e YANAI, 2012	✗	✗	✗	✓	✗	✗	✗	✗	✓
FENG et al., 2014	✗	✗	✗	✓	✗	✗	✗	✓	✓
ILINA et al., 2012	✗	✗	✗	✓	✗	✗	✗	✗	✓
PANIAGUA et al., 2013	✗	✗	✗	✓	✓	✗	✓	✓	✓
GAO et al., 2013	✗	✗	✗	✓	✓	✗	✓	✗	✓
GRABOVITCH-ZUYEV et al., 2014	✗	✗	✗	✓	✗	✓	✗	✗	✗
DAO et al., 2013	✗	✗	✗	✓	✗	✗	✗	✓	✓
PSALLIDAS et al., 2013	✗	✗	✗	✓	✓	✗	✓	✗	✓
CHOUDHURY et al., 2012	✗	✗	✗	✓	✗	✗	✗	✗	✓
BECKER et al., 2011	✗	✗	✗	✓	✓	✓	✗	✓	✓
BECKER et al., 2012	✗	✗	✗	✓	✓	✓	✗	✗	✓
BRENNER e IZQUIERDO 2013	✗	✗	✗	✓	✗	✓	✗	✗	✓
PG++ (este trabalho)	✓	✓	✓	✓	✓	✓	✓	✓	✓

3.4 Considerações finais

Este capítulo apresentou os principais trabalhos relacionados com a proposta desta dissertação. Relacionando-se os principais trabalhos acadêmicos da atualidade, promoveu-se uma discussão sobre as abordagens existentes para anotação de localização e eventos. Conforme apresentado na tabela 3.1, fica claro que os trabalhos estudados não abordam a anotação de localização e eventos sociais ao mesmo tempo. Com isso, surge a necessidade de uma abordagem que englobe todas as características elencadas, considerando-se o contexto de coleções de fotografias pessoais.

Diante desse cenário, este trabalho difere das abordagens estudadas, por apresentar a solução da anotação de localização em conjunto com a anotação de eventos. Em relação à anotação de localização, propõe-se uma abordagem para realizar a anotação de localização de maneira automática e autoadaptável às mudanças no comportamento do usuário. Dessa forma, a anotação de localização pode extrair o melhor das técnicas existentes, aumentando a cobertura de propagações corretas. Em relação à anotação de eventos sociais, propõe-se uma anotação automática que realiza não só a anotação em si, mas também o agrupamento de fotografias de um mesmo evento.

Capítulo 4 – Anotação de localização

A informação da localização geográfica das fotografias pode ser usada para aperfeiçoar a organização e recuperação delas. O uso de dispositivos *GPS* permite fácil obtenção da localização de fotografias, porém falhas no receptor geram um trabalho adicional para o usuário. Como corrigir as falhas ocasionadas pelo receptor *GPS* manualmente é uma tarefa trabalhosa, algumas técnicas de predição de localização e de propagação de localização foram desenvolvidas.

Na literatura estudada, constata-se que alguns trabalhos fazem uso de processamento digital de imagens para encontrar a localização de uma determinada fotografia. Esse processamento é usado para detectar objetos repetidos, pessoas e similaridades visuais entre fotografias, por exemplo. A propagação de localização ocorre de acordo com a semelhança entre as fotografias baseada nas informações coletadas pelo processamento digital de imagens.

O avanço das câmeras fez com que o processamento digital de imagens pudesse ser inserido diretamente nas máquinas que capturam as fotografias (detecção de faces, por exemplo). Dessa forma, com maiores avanços na tecnologia de captura de fotografias, talvez as informações coletadas com o processamento digital de imagens possam ser associadas diretamente à fotografia.

O processamento de imagens através de técnicas de visão computacional resulta em informações sobre a imagem, tais como pessoas presentes, se a captura ocorreu em ambiente aberto ou fechado. Essas informações servem para descrever o contexto no qual a fotografia se encontra. A evolução das câmeras vem inserindo esse processamento digital direto na máquina. Destarte, a propagação de localização pode ser restringida ao uso de informações de contexto para encontrar a posição de uma determinada fotografia.

Para atenuar o problema de anotação de localização presente em coleções pessoais de fotografias digitais, propõem-se, neste trabalho, dois algoritmos de propagação automática de localização: um baseado na trajetória do usuário e outro

baseado nos eventos pessoais do usuário. Tais algoritmos fazem uso das informações de contexto das fotografias e não realizam processamento digital de imagens.

O primeiro algoritmo pode ser útil para a anotação de coleções de fotografias de usuários que se locomovem bastante ou que estiveram em uma viagem incluindo passeios pelo local visitado. O segundo pode ser eficiente para fotografias pertencentes a um mesmo evento ou para grupos de usuários que têm eventos pessoais em comum. Dessa forma, os dois algoritmos resolvem o mesmo problema para cenários específicos.

A resolução de um mesmo problema de maneiras diferentes está cada vez mais comum. Resultante disso, surgem diferentes soluções, cada vez mais específicas, que funcionam melhor para diferentes cenários do problema. Com a proposição dessas diferentes soluções, advém a dificuldade de identificar os cenários em que cada técnica apresenta melhor resultado. As bases de dados estão cada vez maiores e o crescimento delas está ficando mais rápido. Logo, escolher manualmente as melhores técnicas de acordo com a base de dados é inviável.

Além do rápido crescimento das bases de dados, o comportamento dos usuários é imprevisível. Essa mudança de comportamento pode acarretar uma mudança na qualidade dos resultados de qualquer técnica de anotação de localização que esteja sendo executada. Como a maioria das bases de dados de fotografias é grande e o comportamento do usuário pode mudar inesperadamente, a identificação manual da técnica que melhor se aplica é inviável.

Muitas técnicas de propagação de localização geográfica foram propostas (GONG et al., 2011; GAO et al., 2012; HAYS e EFROS, 2008; HOLLENSTEIN e PURVES, 2013; LACERDA et al., 2013; IVANOV et al., 2012) para resolver o problema da anotação de localização em fotografias digitais. Essas técnicas, conforme visto na seção anterior, apresentam resultados diferentes de acordo com o cenário de teste (detalhados no item C da seção 4.1.5) a que são submetidas. Dessa forma, é necessário escolher a técnica de anotação de localização automaticamente.

Neste capítulo, é proposta uma seleção automática que faz uso de métodos de inteligência artificial para selecionar a melhor técnica, dentre as soluções descritas neste capítulo. A principal contribuição da seleção automática proposta é minimizar o

erro gerado por outras técnicas, fazendo uso de modelos matemáticos que auxiliem a escolha da alternativa que melhor satisfaça a propagação de localização.

Como visto no estudo exposto na seção anterior, cada coleção de fotografias conta com uma ou mais técnicas que apresentam melhores resultados que as demais. Tomando-se como exemplo um grupo de usuários que se locomovem muito e não possuem relacionamentos entre si, o algoritmo baseado na trajetória dos usuários tende a ser o mais adequado. Um grupo de usuários presentes nos eventos pessoais uns dos outros deve se enquadrar melhor no algoritmo baseado em eventos pessoais. Assim, o algoritmo de seleção automática de técnica de propagação de localização permite que essas características individuais de cada grupo de usuário sejam alcançadas e que erros futuros sejam evitados.

Visando à necessidade de diminuição no trabalho do usuário em anotar a localização geográfica em coleções de fotografias pessoais, as principais contribuições apresentadas neste capítulo são:

1. O algoritmo de propagação de localização baseado em eventos compartilhados: usa os eventos pessoais das pessoas presentes nas fotografias;
2. O algoritmo de propagação de localização baseado na trajetória: usa as informações das trajetórias das pessoas presentes nas fotografias do usuário;
3. Uma análise comparativa entre os algoritmos propostos e duas técnicas de propagação existentes na literatura: Agrupamento Temporal e Correlação Social. A primeira é a reprodução do trabalho de Lacerda et al. (2013) e a segunda, a reprodução do trabalho de Gong et al. (2011). Esses dois trabalhos da literatura foram escolhidos, porque foi possível replicar a técnica de maneira fiel. A comparação é feita por meio de experimentos, considerando-se as seguintes métricas: precisão, acurácia e cobertura. Ao melhor do nosso conhecimento, uma análise como essa ainda não foi feita na literatura. Os experimentos indicam que as técnicas estudadas apresentam bons resultados para cenários diferentes. Dessa forma, a melhor técnica de propagação dependerá da natureza dos dados e do perfil do usuário;
4. Escolha automática da técnica de propagação;
5. Abordagem autoadaptável ao perfil da base de dados e dos usuários individualmente;

6. Minimização do erro de propagação na coleção de fotografias dos usuários.

4.1 Propagação de localização

Nesta seção, os algoritmos de propagação de localização propostos são explicados. As fotografias utilizadas no experimento possuem as seguintes informações para uso das técnicas: data e hora, localização geográfica e pessoas anotadas. Para todas as técnicas que serão explicadas, considere:

- **u**: um usuário qualquer;
- **F^u**: o conjunto de todas as fotografias de um usuário u;
- **FC^u**: o conjunto de todas as fotografias que não pertencem a **F^u** e que o usuário u aparece;
- **f_i^u**: uma fotografia pertencente ao conjunto **F^u** representada por uma tupla na forma <geo, h, d, dh>, em que "geo" é o ponto de latitude e longitude da fotografia, "h" é a hora do dia, "d" é o dia da semana e "dh" é a data com o horário exato em que a fotografia foi tirada;
- **fc_i^u**: semelhante a **f_i^u**, em que **fc_i^u** pertence a **FC^u**;
- **dif(d1,d2)**: a função que calcula a diferença em horas entre duas datas;
- **LT**: limiar de tempo que tem o valor de 24 horas. Foi adotado o valor arbitrário de 24 horas em razão das conclusões de trabalhos na literatura que analisaram a maneira como os usuários organizam suas fotografias (RODDEN e WOOD, 2003; NAAMAN e NAIR, 2008; COOPER et al. 2005);
- **LS**: limiar de similaridade entre eventos com o valor de 60%. O limiar de similaridade adotado é o proposto por Figueirêdo et al. (2012).

A função "centro" calcula o centro de uma geometria ou, de maneira equivalente, o centro de massa de uma geometria como um ponto. Para múltiplos pontos, o cálculo é feito pela média aritmética das coordenadas e entrada. Para múltiplas linhas, calcula-se com uma ponderação pelo comprimento de cada linha. Para múltiplos polígonos, a ponderação é considerada por meio da área dos polígonos.

Essa definição da função que calcula o centroide é a encontrada no PostGIS².

4.1.1 Técnica de eventos compartilhados

O trabalho de Figueirêdo et. al (2012) propõe uma técnica de detecção de eventos em fotografias digitais capturadas por câmeras distintas. A detecção de eventos separa todos os possíveis eventos na coleção de fotografias e calcula um índice de similaridade entre esses eventos.

Define-se:

E : o conjunto de eventos;

$ef_i^u \in E$: o evento que contém a fotografia f_i^u ;

$\forall e_k \in E$ existe um SM_{ik}^u , em que SM_{ik}^u representa a similaridade entre os eventos e_k e ef_i^u .

Este trabalho sugere a utilização desses eventos pessoais encontrados para calcular a localização da fotografia f_i^u , assumindo que fotografias presentes em um mesmo evento estão em um raio próximo. Seja:

EC_i^u um subconjunto de E correlacionado a ef_i^u , tal que $\forall e_k \in EC_i^u \rightarrow SM_{ik}^u \geq LS$;

FC o conjunto de fotografias com localização presente nos eventos de EC , tal que $\forall fc \in FC \rightarrow \text{dif}(\text{dh}(fc), \text{dh}(f_i^u)) \leq LT$;

Considerando-se fc uma fotografia de FC , a técnica procura a fotografia fc^* e propaga a localização de fc^* para f_i^u . A equação 4.1 define fc^* .

$$fc^* = \text{MIN}(\text{dif}(\text{dh}(p), \text{dh}(f_i^u))); \quad (4.1)$$

4.1.2 Técnica de trajetória

O trabalho de Lacerda et. al (2013) propõe o uso da velocidade do usuário para encontrar inconsistências nas localizações das fotografias. Inspirada na ideia de velocidade do usuário, a técnica de trajetória aqui proposta assume que o percurso entre fotografias com localização é uma reta para propagar localização através deste percurso. A técnica não só usa a trajetória do usuário u , mas também a trajetória de

² <http://postgis.org/>

todos os contatos presentes na fotografia f_i^u , que não tem localização. Dessa forma, a ideia de correlação social é inserida na busca da localização da fotografia.

Considere para cada pessoa p presente em f_i^u :

$$FA^p = F^p \cup FC^p; \quad (4.2)$$

É feita uma decomposição de FA^p em dois conjuntos:

1. FA_{geo}^p : o conjunto de todas as fotografias em FA^p que possuem localização;
2. FA_{notGeo}^p : o conjunto de todas as fotografias em FA^p que não possuem localização;

Para propagar a localização geográfica para a fotografia $f_i \in FA_{notGeo}^p$, é preciso encontrar as fotografias f_1^* e f_2^* , em que:

$$f_1, f_2 \in FA_{geo}^p; \quad (4.3)$$

$$dh(f_1) < dh(f_i); \quad (4.4)$$

$$dh(f_2) > dh(f_i); \quad (4.5)$$

$$\text{dif}(dh(f_1), dh(f_i)) \leq LT; \quad (4.6)$$

$$\text{dif}(dh(f_2), dh(f_i)) \leq LT; \quad (4.7)$$

$$f_1^* = \text{MIN}(dh(f_i) - dh(f_1)); \quad (4.8)$$

$$f_2^* = \text{MIN}(dh(f_2) - dh(f_i)); \quad (4.9)$$

Para computar a possível localização geo^p para a fotografia f_i , considerando-s a trajetória de p , usa-se interpolação linear com base em:

$$z = \frac{dh(f_i) - dh(f_1^*)}{dh(f_2^*) - dh(f_1^*)}; \quad (4.10)$$

$$geo^p = f_1^* \cdot (1 - z) + z \cdot f_2^*; \quad (4.11)$$

A interpolação linear não considera a curvatura da Terra, porém, como os algoritmos aplicam-na a pequenas distâncias, o erro gerado é irrelevante. Esse processo é repetido para cada pessoa p em f_i ; então, a localização é calculada através do centroide do conjunto composto por todas as geo^p encontradas.

4.1.3 Exemplos

Nesta subseção, é apresentado um exemplo para cada algoritmo de propagação

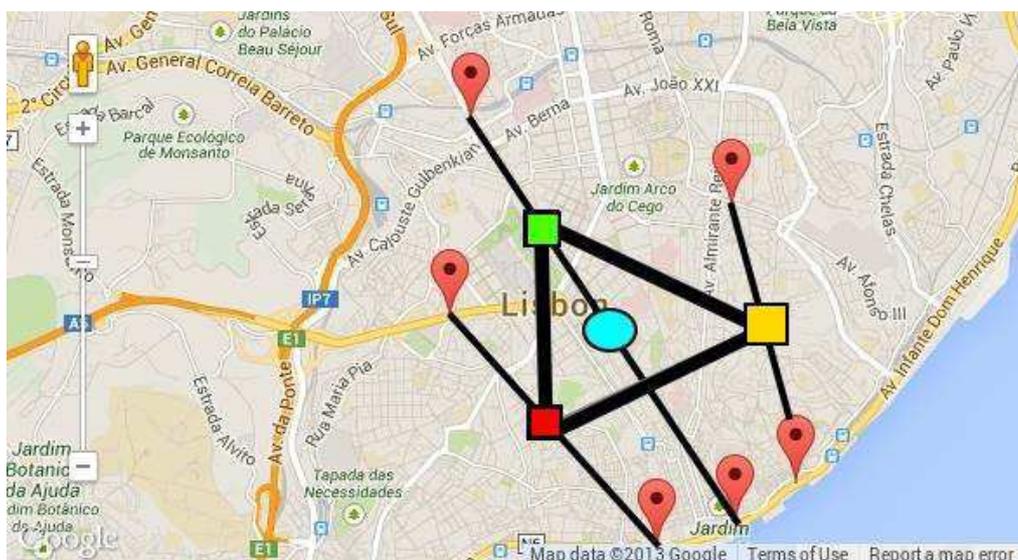
de localização proposto. Primeiro, é exemplificado o algoritmo baseado na trajetória do usuário e, em seguida, é descrito um exemplo do algoritmo baseado nos eventos pessoais dos usuários.

A. Algoritmo baseado na trajetória do usuário

Considera-se que alguns amigos capturaram algumas fotografias e que apenas uma fotografia estaria sem localização geográfica. As pessoas p_1 , p_2 , p_3 e p_4 estão presentes na fotografia f_i^u , fotografia sem localização, mas apenas p_1 , p_2 e p_3 possuem uma trajetória relacionada à fotografia.

Na figura 4.1, são apresentadas as trajetórias de p_1 , p_2 e p_3 como uma linha reta entre os marcadores no mapa. Os vértices do triângulo são as possíveis localizações relacionadas com cada trajetória. O ponto dentro do triângulo seria o centroide de todas as possíveis localizações e esse ponto é a localização propagada.

Figura 4.1. Exemplo com algoritmo baseado na trajetória



B. Algoritmo baseado em eventos pessoais

Considera-se que dois amigos, a_1 e a_2 , participaram de uma festa de carnaval com duração de 5 horas. O primeiro tem o costume de capturar inúmeras fotografias e o segundo prefere capturar apenas os principais momentos da festa. Supõe-se que a_1 esteja presente nas fotografias capturadas por a_2 e que essas fotografias não possuam localização geográfica. Ao final da festa, a_1 registrou a festa a cada 20 minutos e a_2 a cada 1 hora. Ao carregar as fotografias capturadas no sistema, a detecção de eventos pessoais indicou que os eventos de a_1 e a_2 possuem uma similaridade maior que o

F^u , tal que:

$$\left(\bigcup_{j=1}^k g_j \right) = F^u; \quad (4.12)$$

Os grupos possuem fotografias com distâncias temporais menores ou iguais a t_{max} minutos, considerando-se fotografias consecutivas, e podem ser definidos da seguinte forma:

$$f_1 \in g_1; \quad (4.13)$$

$$f_{i+1} \in g_s \text{ if } (t_{i+1} - t_i) \leq t_{max}; \quad (4.14)$$

$$f_{i+1} \in g_{s+1} \text{ if } (t_{i+1} - t_i) > t_{max}; \quad (4.15)$$

Para cada grupo g_j e considerando-se lf_i^u a localização da fotografia f_i^u , a iteração vai procurar as fotografias que não possuem localização, ou seja, lf_i^u é vazio. Para cada fotografia sem localização, procura-se uma fotografia f_j^u com localização que seja mais próxima temporalmente dentro do mesmo grupo de f_i^u e, então, a localização de f_j^u é propagada para f_i^u .

B. Correlação Social

A técnica de Correlação Social é baseada no trabalho de Gong et al. (2011) e utiliza as localizações geográficas de possíveis vizinhos para encontrar a localização do usuário no exato momento em que a fotografia foi tirada.

Antes de encontrar a localização de uma fotografia, é preciso identificar qual pessoa tem a maior correlação social com o usuário que capturou a fotografia f_i^u . Seja Q_i^j a quantidade de fotografias em que o usuário i e o usuário j aparecem. A correlação entre o usuário i e o usuário j é R_i^j , definida pela equação 4.16.

$$R_i^j = \frac{Q_i^j}{Q_i^i}; \quad (4.16)$$

Dessa forma, R_i^j indica a porcentagem de fotografias das quais o usuário j participa em relação às fotografias do usuário i . Seja C^u o conjunto de contatos vizinhos de um usuário u . Levando-se em conta que será sugerida a localização de uma

fotografia f_i^u , deve-se encontrar o contato c^* com a maior correlação social e com pelo menos uma fotografia em um intervalo menor ou igual à LT , considerando-se f_i^u . Tem-se que:

$$\forall c \in C^u \rightarrow R_u^c \leq R_u^{c^*}, \quad (4.17)$$

e

$$\exists f_j^{c^*} \in F^{c^*}, \text{ tal que } dif(dh(F_x^{c^*}), dh(F_i^u)) \leq LT; \quad (4.18)$$

Seja **antes**(v, f_i^u) a função que encontra a última fotografia tirada anterior a f_i^u , em que o vizinho v aparece, e seja **depois**(v, f_i^u) a função que encontra a primeira fotografia tirada após f_i^u , em que o vizinho v aparece. As duas funções procuram uma fotografia em um intervalo menor ou igual à LT em relação à fotografia passada como parâmetro. Caso não exista, as funções retornam um valor vazio. Dessa forma, há a propagação de localização por correlação social para uma fotografia f_i^u , definida como:

$$CS_i^u = \text{Centro}(\text{antes}(C^*, f_i^u), \text{depois}(C^*, f_i^u), \text{antes}(u, f_i^u), \text{depois}(u, f_i^u)); \quad (4.19)$$

4.1.5 Validação

Nesta subseção, expõe-se o método utilizado para coletar os dados, as características sobre a base de fotografias usada para testar as técnicas, as métricas adotadas para comparar as técnicas observadas e, por fim, como as métricas foram validadas para gerar os resultados obtidos.

A. Base de dados

Para este trabalho, foi utilizada uma base de dados de aproximadamente 7.900 fotografias divididas entre 41 usuários. Cada fotografia disponibiliza os seguintes metadados: data e hora em que foi tirada, localização geográfica e pessoas presentes. Apenas 13% das fotografias vieram de câmeras que possuem GPS integrado, logo 87% das localizações observadas para as câmeras que não possuem GPS foram indicadas pelos próprios donos das fotografias.

Essa base de dados foi extraída do PhotoGeo (FIGUEIREDO et al. 2012), uma plataforma para gerenciamento de coleções de fotografias pessoais com foco em anotação de pessoas e eventos pessoais.

B. Métricas

As métricas adotadas foram precisão, cobertura e acurácia. Cada uma foi definida na seção 2.1.4 do capítulo de fundamentação teórica. Todas essas métricas possuem importâncias individuais e devem ser analisadas tanto separadamente como em conjunto. A métrica nomeada cobertura indica o número de fotografias que recebeu a localização dentro do limiar de erro adotado, considerando-se todo o conjunto de fotografias.

Existem algumas fotografias cuja localização os algoritmos de propagação não conseguem encontrar. A métrica denominada precisão não leva em consideração essas fotografias que não receberam localização e aponta o quão correta é a propagação de um determinado algoritmo. A métrica chamada acurácia caracteriza a porcentagem de fotografias que não receberam uma localização além do limiar de erro.

C. Cenários de teste

Nesta subseção, os cenários de teste considerados na validação dos algoritmos são explicados. Eles se diferenciam por meio da construção do conjunto de treinamento a ser usado pelos algoritmos. Essa escolha pode ocorrer de três maneiras: por fotografia, por usuário e por evento.

- **Por fotografia:**

Considerando-se o conjunto de fotografias F , contendo todas as fotografias do sistema, o conjunto de treinamento TS é construído com base na escolha aleatória de fotografias até que ele tenha um número de fotografias igual a 50% do número de fotografias do conjunto F .

Esse cenário de teste representa os sistemas que incluem usuários com câmeras que possuem um dispositivo para capturar a localização da fotografia. A escolha aleatória das fotografias para treinamento representa os dispositivos que, por alguma razão (mau funcionamento ou dispositivo de localização desativado, por exemplo), não armazenaram a localização de algumas fotografias.

- **Por usuário:**

Considerando-se o conjunto contendo todos os usuários U e F^u o conjunto das fotografias de um usuário "u" pertencente a U , o conjunto de treinamento TS é formado pela inserção gradual das fotografias de usuários aleatórios pertencentes a U . Dessa forma, uma nova inserção de um F^u aleatório em TS só ocorrerá se o tamanho de TS for menor que 50% do tamanho do conjunto de fotografias F .

Esse cenário de teste corresponde aos sistemas que possuem alguns usuários que armazenam e outros que não armazenam a localização nas fotografias. Como o uso de dispositivos GPS consome energia, alguns usuários preferem desabilitar esse recurso do dispositivo móvel. A escolha aleatória dos usuários que terão as fotografias no conjunto de treinamento representa os usuários que armazenam a localização nas fotografias.

- **Por evento:**

Considerando-se o conjunto de todos os eventos cadastrados no sistema E e sendo FE^e o conjunto de fotografias de um evento "e" pertencente a E , a construção do conjunto de treinamento TS se dá de maneira análoga à descrita no cenário por usuário. As inserções graduais em TS ocorrem por meio da escolha de um FE^e aleatório, até que TS possua um número de fotografias igual ou superior a 50% da quantidade de fotografias de F .

Esse cenário de teste representa a possibilidade de mau funcionamento em eventos localizados em área não alcançada satisfatoriamente pelo sinal de sistemas de localização. Dessa forma, o conjunto de treinamento corresponderá aos eventos em que os dispositivos conseguiram pegar a localização geográfica.

D. Experimentos

Uma validação dos resultados é necessária para garantir boa representatividade e também boa significância estatística. O conjunto de treinamento usado foi de 50% das fotografias presentes na base de dados, logo a outra metade foi usada para comparar a localização propagada com a localização original. Para cada réplica (num total de 30 réplicas), o conjunto de treinamento foi escolhido usando randomização, para garantir que o resultado seja imparcial. No algoritmo 4.1, a coleta dos resultados é detalhada.

Algoritmo 4.1. Coleta dos resultados

```
for (replica = 0; replica < 30; replica++) {  
    fotografiasRemoverTemp = getFotografiasARemover();  
    for(Tecnica tec in tecnicas){  
        removerTemporariamente(fotografiasRemoverTemp);  
        for(fotografia in fotografiasRemoverTemp){  
            localizacaoSugerida = tec.sugerirLocalizacao(fotografia);  
            salvarSugestao(localizacaoSugerida);  
        }  
        restaurarLocalizacaoOriginal(fotografiasRemoverTemp);  
    }  
}
```

Uma sugestão é considerada correta quando a sua distância para a localização original é menor ou igual ao limiar de erro adotado, que, neste trabalho, foi 100 metros. Utilizou-se o menor limiar encontrado na literatura, porque, se uma técnica de propagação acerta uma propagação para uma fotografia f com limiar espacial de 100 metros, essa mesma técnica também acertará a propagação para a fotografia f em qualquer limiar maior que 100 metros. Dessa forma, o uso do limiar de erro de 100 metros é favorável aos resultados obtidos, pois considera a realidade de vários sistemas que necessitam de propagação de localização.

As fotografias que terão as localizações removidas temporariamente são as que estão fora do conjunto de treinamento. Antes de cada técnica ser executada, as localizações das fotografias (que não estão no conjunto de treinamento) são removidas temporariamente e, logo após, é feita a tentativa de propagação. Com todas as sugestões de todas as réplicas armazenadas, podem ser analisadas as métricas propostas.

Ao final da coleta, fez-se uma observação dos intervalos de confiança de cada técnica em relação às três métricas analisadas. Considerando-se um nível de significância $\alpha = 5\%$, os intervalos de confiança garantem 95% de confiança nos resultados, dispensando o uso de algum teste estatístico mais elaborado, como teste Student-T ou Wilcoxon (BOSLAUGH e WATTERS, 2008). Dessa forma, pode-se chegar a conclusões em relação à população e não somente para a amostra do experimento.

4.1.6 Resultados

Nesta seção, apresentam-se os resultados obtidos para cada técnica em relação a todas as métricas observadas e, em seguida, discutem-se tais resultados. Os resultados encontrados seguiram o mesmo comportamento para todas as métricas observadas. Dessa forma, é possível criar um único *ranking* de técnicas. As técnicas são representadas por siglas:

- **SC:** Correlação Social;
- **SE:** Eventos Compartilhados;
- **TT:** Trajetória;
- **TC:** Agrupamento Temporal.

Existem dois momentos em que as fotografias são escolhidas aleatoriamente:

1. Quais fotografias farão parte do conjunto de treinamento?
2. Qual a próxima fotografia sem localização que será analisada pela técnica?

O primeiro momento destina-se a garantir que o conjunto de treinamento não seja tendencioso. Escolher as partes que farão parte do conjunto de treinamento aleatoriamente é fundamental para garantir a representatividade dos resultados obtidos. O segundo momento objetiva melhorar o desempenho das técnicas.

A propagação de localização para uma fotografia pode interferir na propagação de localização para outra fotografia, pois, se a primeira propagação estiver errada, poderá repassar o erro para as demais fotografias consideradas semelhantes pelos algoritmos. Para todas as métricas observadas, houve diferenças entre 10% e 20% em favor da propagação de localização em ordem aleatória.

Dessa forma, apresentam-se os resultados de cada métrica, considerando-se uma sequência aleatória de propagação. Nas figuras 4.3, 4.4 e 4.5, são mostrados os resultados de cada técnica para as três métricas observadas. Os valores apresentados nas figuras são as medianas dos conjuntos de resultados de cada técnica, e a comparação entre as medianas corresponde aos resultados obtidos na comparação feita pelos testes estatísticos realizados.

Figura 4.3. Acurácia das Técnicas

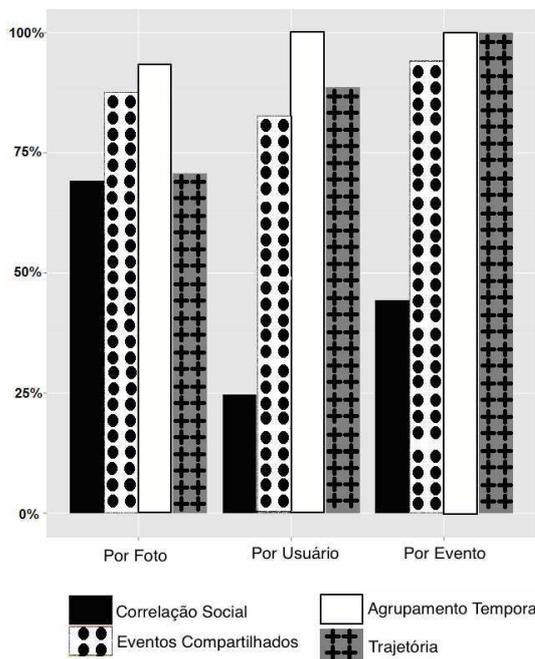


Figura 4.4. Cobertura das Técnicas

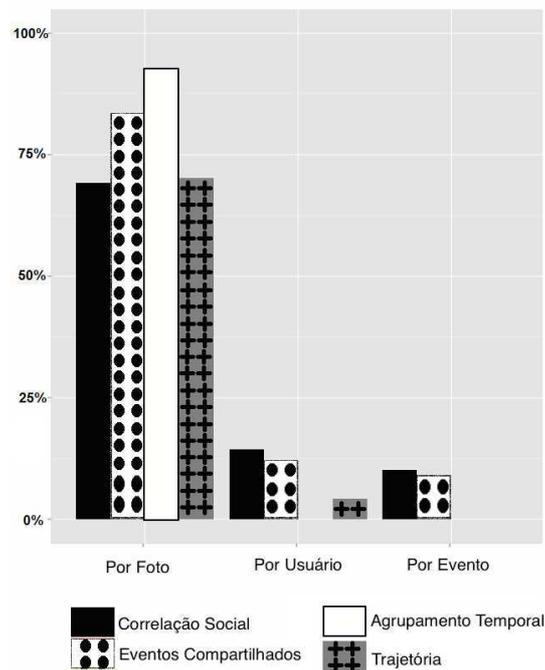
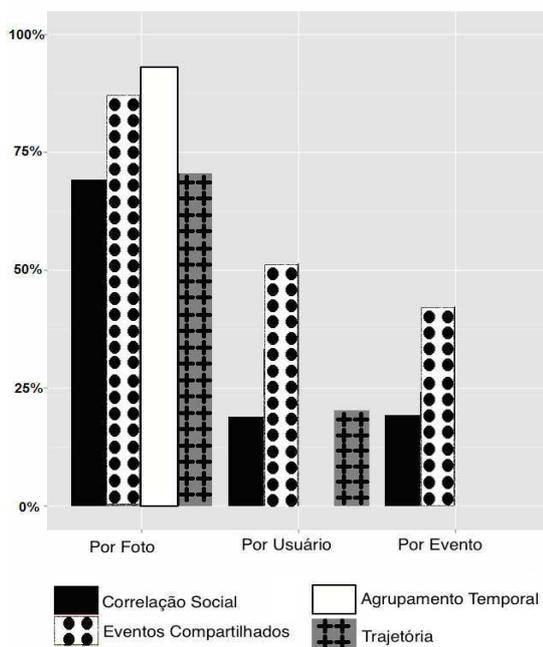


Figura 4.5. Precisão das Técnicas



Constata-se que a técnica de Agrupamento Temporal (usando um t_{max} de 24 horas) para a realidade de propagação de localização automática é a melhor em relação à precisão, cobertura e acurácia. Os resultados obtidos foram bons para todas as

métricas, fazendo com que a propagação possa ser usada em vários sistemas.

A. Abordagem por fotografia:

A abordagem construindo o conjunto de treinamento pela escolha aleatória de fotografias, sem considerar os usuários ou eventos, apresentou o mesmo *ranking* de algoritmos para todas as métricas. Esse *ranking* de algoritmos é o seguinte:

$$SC < TT < SE < TC$$

B. Abordagem por usuário:

O cenário de teste que considera a construção do conjunto de treinamento pela escolha aleatória de usuário apresentou diferentes *rankings* para as métricas. São eles:

- Precisão: $TC < TT = SC < SE$;
- Cobertura: $TC < TT < SC = SE$;
- Acurácia: $SC < SE < TT < TC$.

A acurácia apresentou uma ordem inversa, pois os algoritmos com a cobertura baixa possuem alto número de fotografias sem localização (S_n é grande). Com isso, automaticamente o valor da acurácia aumenta.

C. Abordagem por evento:

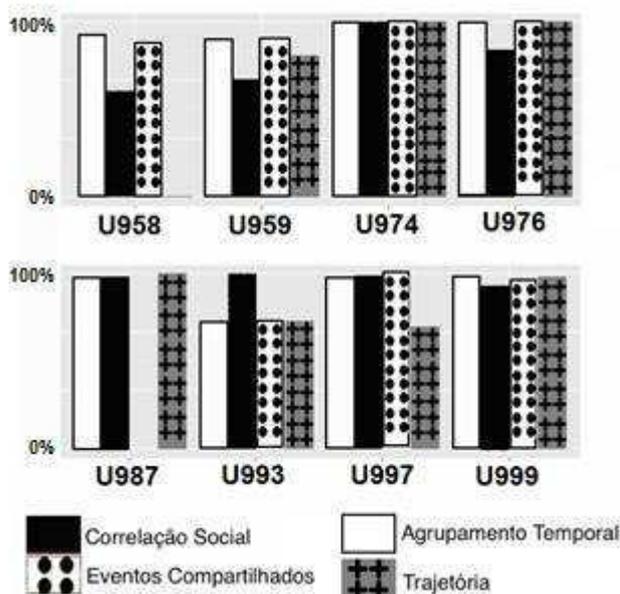
O cenário de teste que monta o conjunto de treinamento considerando os eventos dos usuários apresentou resultados similares aos do cenário de teste considerando usuários aleatórios. Destarte, os *rankings* são os mesmos para todas as métricas.

D. Análise das técnicas por usuário:

Por fim, foi feita, separadamente, uma análise do comportamento de cada um dos algoritmos em relação aos usuários da base de dados. Os resultados indicaram que o comportamento de cada algoritmo varia conforme a coleção de fotografias de cada usuário. Na figura 4.6, são mostrados oito usuários para os quais os algoritmos apresentaram diferenças entre os usuários para a métrica de precisão. Esses resultados representam o cenário de teste de escolha aleatória de fotografias, porém, nos demais

cenários, essa variação por usuário também foi detectada. Na figura 4.6, cada grupo de quatro colunas representa um usuário e cada coluna representa um algoritmo.

Figura 4.6. Comportamento por Usuário



4.2 Seleção automática de técnica de propagação

Nesta seção, a seleção automática de técnicas de propagação de localização é apresentada e detalhada. Os estudos mostrados na seção anterior indicam que a melhor técnica pode ser diferente para cada usuário. Dessa forma, é preciso identificar automaticamente a melhor técnica para cada um desses usuários. Para que ocorra essa identificação, existem três alternativas:

1. Definir características para o cenário adequado de cada técnica e estabelecer rotinas condicionais que executarão a técnica adequada baseada nas características pré-definidas;
2. Unir as técnicas de propagação existentes em uma só;
3. Usar aprendizado de máquina para que os valores dessas características sejam identificados automaticamente para cada técnica de propagação, considerando-se a individualidade de cada usuário.

A primeira alternativa é inviável, pois não existem dados para garantir que os cenários ficarão com suas características fixas. Com isso, o crescimento da base de

dados pode fazer com que uma determinada técnica supere outra em algum dos cenários preestabelecidos. Em virtude dessa limitação, a primeira alternativa é descartada.

A segunda alternativa não é praticável, pois as técnicas de propagação são bastante distintas e unir todas em uma única técnica levaria ao mesmo problema da alternativa anterior. Como as técnicas são muito diferentes, a união delas se daria por meio da identificação de características das fotografias para executar o algoritmo de propagação mais adequado. Dessa forma, o problema apontado na primeira alternativa se repete, porém, na primeira alternativa, ele ocorre no nível da base de dados e não no nível de fotografia como na segunda alternativa.

Em razão das limitações das duas primeiras alternativas, a terceira foi usada para realizar a seleção automática da técnica adequada para cada cenário identificado na seção anterior, considerando-se a individualidade de cada usuário. Por se tratar de uma alternativa com aprendizado de máquina, é preciso fornecer algum conjunto de treinamento para que a seleção automática possa identificar a melhor técnica.

Ressalta-se que alguns metadados envolvendo as fotografias foram levantados para o aprendizado de máquina. Esses metadados são reunidos em uma tabela com seus respectivos valores normalizados. Além dos metadados, cada uma das técnicas usadas na seleção deve conter uma coluna na tabela indicando se a respectiva técnica acertou ou não a localização da fotografia. Cada coluna da tabela citada é descrita na tabela 4.1.

A normalização dos metadados depende de cada atributo, podendo ser normalizado pelas funções N1 ou N2, descritas a seguir:

$$N1(x, \max) = \left(\frac{x}{\max}\right) \quad (4.23)$$

$$N2(x, \max) = 1 - \left(\frac{x}{\max}\right) \quad (4.24)$$

Tabela 4.1. Dicionário de dados referentes à tabela de metadados.

#	Atributo	Tipo	Comentário
1	ID da fotografia	Inteiro	Armazena o id da fotografia
2	ID do usuário	Inteiro	Armazena o id do usuário dono da fotografia
3	Antes	Ponto flutuante	Valor normalizado da distância em segundos da última fotografia com localização
4	Depois	Ponto flutuante	Valor normalizado da distância em segundos da fotografia seguinte com localização
5	Quantidade de pessoas	Ponto flutuante	Valor normalizado da quantidade de pessoas presentes na fotografia
6	Fotos dos contatos	Ponto flutuante	Valor normalizado da quantidade de fotografias com localização dos contatos presentes na fotografia
7	Fotos evento	Ponto flutuante	Valor normalizado da quantidade de fotografias com localização presente no mesmo evento
8	Fotos álbum	Ponto flutuante	Valor normalizado da quantidade de fotografias com localização presentes no mesmo álbum
9	Fotos usuário	Ponto flutuante	Valor normalizado da quantidade de fotografias com localização pertencentes ao mesmo usuário
10	Distância entre "antes" e "depois"	Ponto flutuante	Valor normalizado da distância temporal em segundos das fotografias com localização encontradas pelos atributos "antes" e "depois"
11	Fotos no mesmo dia	Ponto flutuante	Valor normalizado da quantidade de fotografias com localização capturadas no mesmo dia da fotografia sem localização
12	Técnica 1	Inteiro	Valor 0 ou 1, indicando se determinada técnica gerou o melhor resultado (1) ou não (0)
13	Técnica 2	Inteiro	Valor 0 ou 1, indicando se determinada técnica gerou o melhor resultado (1) ou não (0)
14	Técnica 3	Inteiro	Valor 0 ou 1, indicando se determinada técnica gerou o melhor resultado (1) ou não (0)
15	Técnica 4	Inteiro	Valor 0 ou 1, indicando se determinada técnica gerou o melhor resultado (1) ou não (0)
...
X	Técnica N	Inteiro	Valor 0 ou 1, indicando se determinada técnica gerou o melhor resultado (1) ou não (0)

O processamento para encontrar os valores dos atributos é descrito a seguir:

- "ID do usuário" e "ID da fotografia": recebem o valor bruto, sem nenhuma modificação, não há necessidade de normalização;
- "Antes", "Depois" e "Distância entre 'antes' e 'depois'": recebem o valor normalizado a partir da função N2, cujo parâmetro "max" indica o maior valor encontrado no conjunto de valores da coluna em questão antes da normalização;
- "Fotos dos contatos", "Quantidade de pessoas", "Fotos do mesmo evento", "Fotos álbum", "Fotos usuário" e "Fotos no mesmo dia": recebem o valor normalizado a partir da função N1, cujo parâmetro "max" indica o maior valor encontrado no conjunto de valores da coluna em questão antes da normalização;

Para cada usuário, é construída uma tabela listando suas fotografias, os metadados e seus respectivos valores. Dessa forma, o aprendizado de máquina pode identificar quais atributos são importantes para escolher a melhor técnica de propagação e, então, prever a melhor técnica. A tabela de metadados é por usuário, porém a seleção de melhor técnica se dá por fotografia. Com isso, o aprendizado de máquina usa a individualidade de cada usuário para escolher a melhor técnica para cada fotografia, considerando os metadados levantados. Um exemplo de tabela de metadados filtrada para um usuário específico pode ser vista Na figura 4.7. Cada atributo a_i envolvendo as fotografias é representado por uma coluna azul e cada técnica t_j é representada por uma coluna amarela.

Essa abordagem, levando-se em conta também o contexto em que a fotografia se encontra, faz com que a seleção automática considere não só uma única técnica por usuário; logo, fotografias de um mesmo usuário podem receber indicações de técnicas de propagação distintas.

Figura 4.7. Exemplo de tabela de um usuário aleatório

id	a1	a2	a3	a4	a5	a6	a7	...	an	t1	t2	t3	...	tn
1	0	0	0	0.3	0	1	0	...	0	0	1	0	...	1
2	0.3	0.2	0.1	1	0.7	0.9	1	...	1	1	1	1	...	1
3	0	0	0	1	0	0.8	0	...	0	0	1	0	...	1
4	0.4	0	0.5	0	0.3	0	1	...	0	1	0	1	...	0
5	0.9	0.7	0	1	0	1	0	...	0	0	1	0	...	1
...
n	1	1	0	1	0.3	0	0	...	0	0	1	1	...	1

Para realizar o aprendizado de máquina, foram consideradas algumas técnicas de inteligência artificial: fatoração de matrizes, regressão linear múltipla e SVM. A escolha da técnica de aprendizado foi realizada por meio dos resultados obtidos em testes iniciais. Foi criada uma tabela de metadados aleatoriamente e executado o aprendizado com as três técnicas. A técnica que apresentou o menor erro foi escolhida.

A ferramenta Weka³ foi usada para verificar o erro no aprendizado. As técnicas de fatoração de matrizes e SVM apresentaram resultados semelhantes, porém não foram bons. O uso de fatoração de matrizes ou SVM apresentaria resultados piores que a técnica de propagação que faz uso de Agrupamento Temporal. Dessa forma, essas duas técnicas foram descartadas, pois seria melhor usar uma técnica de propagação isoladamente do que fazer a seleção automática por intermédio do aprendizado de máquina com fatoração de matrizes ou SVM.

A técnica de regressão linear apresentou resultados diferentes de fatoração de matrizes e SVM. A regressão linear, com os testes iniciais, conseguiu melhorar a quantidade de fotografias com propagação correta. O uso de regressão linear revelou melhores resultados que as técnicas de propagação individualmente, logo o uso de regressão linear para realizar a seleção automática de técnicas de propagação mostrou indícios de que seria vantajoso. Dessa forma, a técnica de regressão linear múltipla foi escolhida para realizar a seleção automática de técnicas de propagação.

³ <http://www.cs.waikato.ac.nz/ml/weka/>

4.2.1 Modelo de regressão linear múltipla

Nesta seção, a seleção automática de técnicas de propagação de localização, tendo com base a regressão linear, é apresentada e detalhada. O modelo da técnica proposta neste trabalho é exposto na figura 4.8. Os usuários tiram fotografias que vão para seleção automática de técnicas de propagação de localização. A seleção automática possui N técnicas e uma delas será escolhida de acordo com o perfil de cada usuário. Os metadados da nova fotografia são identificados pela seleção automática e passados para a regressão linear múltipla do usuário dono da fotografia. A regressão linear, por meio de uma análise prévia das fotografias com localização, indicará que técnica de propagação deve ser usada para a nova fotografia. A técnica de propagação é executada e a localização é anotada na nova fotografia.

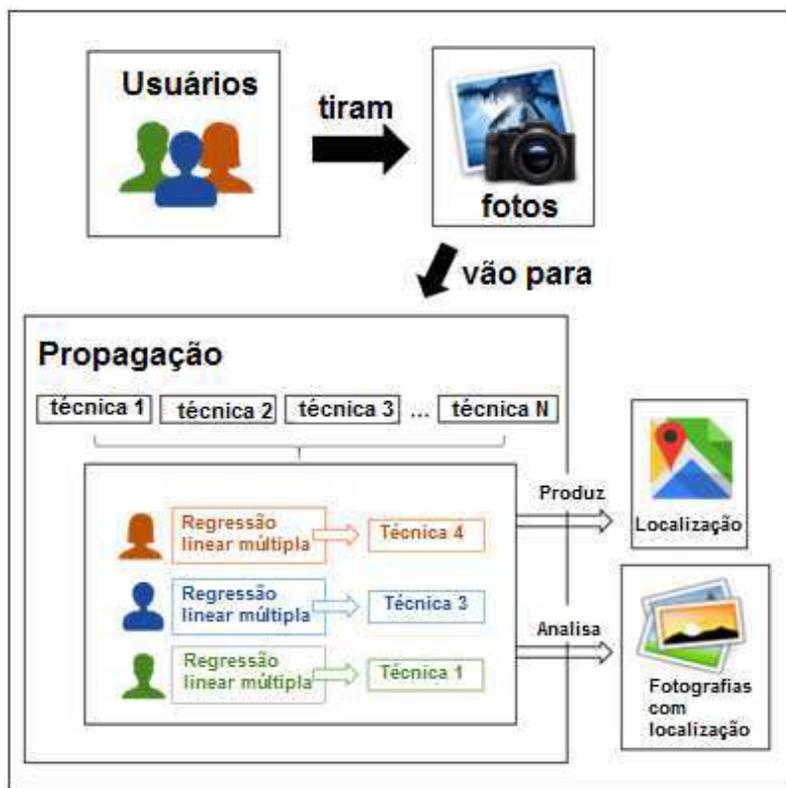
Os resultados apresentados na seção anterior são frutos de uma comparação entre quatro diferentes técnicas de propagação de localização geográfica, sendo duas delas propostas por este trabalho. Os resultados apontam que a escolha da técnica de propagação de localização depende do perfil da base de dados e do perfil do usuário dono das fotografias. Assim, é necessário escolher uma técnica por fotografia, considerando-se o perfil de cada um dos usuários individualmente.

Este trabalho propõe uma seleção automática que identifica qual das técnicas deve ser usada. Essa seleção é feita por meio de regressão linear múltipla, em que cada usuário tem seus modelos de regressão linear para prever que técnica gerará o menor erro e para usá-la na propagação de uma fotografia específica.

As regressões lineares múltiplas usadas neste trabalho podem ser interpretadas como funções que recebem muitas variáveis de entrada e retornam um valor entre 0 e 1, em que 0 indica que a técnica não deve ser usada e 1, que a técnica deve ser usada. A equação a seguir formaliza a regressão linear múltipla usada aqui, sendo que a_n representa o atributo de número n da tabela de metadados, relacionados com as fotografias, descrita na seção anterior, e c_n representa o coeficiente do atributo a_n .

$$rf(a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}) = c_3 \cdot a_3 + c_4 \cdot a_4 + c_5 \cdot a_5 + c_6 \cdot a_6 + c_7 \cdot a_7 + c_8 \cdot a_8 + c_9 \cdot a_9 + c_{10} \cdot a_{10} + c_{11} \cdot a_{11} \quad (4.25)$$

Figura 4.8. Modelo da seleção automática proposta neste trabalho



Para que a regressão linear se processe, é necessário haver um conjunto de treinamento a fim de que os coeficientes c_n possam ser calculados. Com isso, a tabela de metadados associados às fotografias (que receberam a localização por meio de propagação), detalhada na seção anterior, é usada para que as regressões lineares sejam criadas.

Com a tabela construída, as linhas de cada usuário devem ser separadas para que os modelos lineares sejam criados, considerando-se apenas as fotografias dos usuários. As regressões são criadas com base na tabela de metadados relacionados com as fotografias dos usuários. Cada técnica presente na tabela terá sua respectiva regressão linear para indicar se a técnica é adequada ou não para cada nova fotografia. Ao final, tem-se uma regressão $r f_t^u(a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11})$ para o usuário u e técnica de propagação t . Dessa forma, cada usuário u terá uma regressão linear múltipla para cada técnica t .

A propagação deve ocorrer com uma das técnicas da seleção automática, porém os modelos lineares de cada usuário servirão para prever qual delas possivelmente gerará o menor erro. Para cada fotografia sem localização de um usuário u , os atributos

a_3 até a_{11} devem ser encontrados e colocados nas regressões rf_t^u para analisar os resultados das regressões lineares. A técnica a ser usada será aquela relacionada à regressão linear cujo valor foi maior. No caso de empate, a técnica com maior número de acertos até o momento deverá ser usada.

4.2.2 Validação

Nesta subseção, expõe-se como a validação da seleção automática foi realizada. No primeiro tópico, são detalhados o método utilizado para coletar os dados e as características sobre a base de fotografias usada para testar a seleção automática de técnicas de propagação. Ainda no primeiro tópico, a métrica adotada para comparar a seleção automática com a melhor técnica para cada cenário também é especificada. Por fim, no segundo tópico, é explicada a maneira como os experimentos foram realizados para validar a métrica.

A. Base de dados, cenários de teste e métrica

Para a validação da seleção automática de técnicas de propagação, foi utilizada a mesma base de dados descrita na seção anterior. Os cenários de teste também são os descritos no capítulo anterior e são empregados para diferenciar a construção do conjunto de treinamento a ser usado pelas técnicas de propagação. Essa escolha pode ocorrer de três maneiras: por fotografia, por usuário e por evento.

A métrica escolhida foi cobertura (do inglês: *recall*), pois o objetivo da seleção automática de técnicas de propagação de localização é aumentar a quantidade de fotografias que recebem uma localização correta. Seja:

S_c = Quantidade de propagações corretas e

T_f = Total de fotografias sem localização geográfica.

A métrica de cobertura é definida como:

$$\text{Cobertura: } \frac{S_c}{T_f} \quad (4.26)$$

B. Experimentos

Nesta subseção, expõe-se a forma como os resultados foram obtidos e validados. Para coletar os resultados, foi usada a base de dados previamente apresentada, tendo sido considerados apenas os usuários com um número mínimo de dez fotografias em sua coleção. Como as técnicas de propagação e a regressão

necessitam de conjunto de treinamento e os conjuntos não são os mesmos, foi estabelecido um número mínimo de dez fotografias por usuário para que todo o processo de treinamento se desse corretamente.

A coleta dos resultados foi dividida nas seguintes etapas:

1. Cálculo do erro na propagação de localização das técnicas usadas na regressão;
2. Construção da tabela a ser usada na regressão;
3. Teste dos modelos lineares.

- **Etapa 1 - Cálculo do erro na propagação de localização das técnicas usadas na regressão**

A primeira etapa da coleta de resultados é representada pelo pseudocódigo apresentado no quadro 4.1. Para todos os usuários existentes no sistema, metade das fotografias terá sua localização geográfica removida temporariamente e o restante servirá como conjunto de treinamento para que as técnicas façam a sugestão de localização. A escolha das fotografias que terão a localização removida temporariamente é feita aleatoriamente.

As técnicas presentes na seleção automática fazem a propagação de localização e, ao final do experimento, o erro das técnicas é calculado com a função de cálculo de distância geodésica entre a coordenada geográfica sugerida e a coordenada original da fotografia. A primeira etapa é concluída com o erro de cada técnica relacionado às fotografias com localizações removidas temporariamente.

- **Etapa 2 - Construção da tabela de atributos a ser usada na regressão**

A segunda etapa da coleta dos dados consiste na preparação para aplicação da técnica de regressão linear múltipla. Para que a regressão se processe, é necessário montar a tabela com os metadados associados às fotografias e às colunas, indicando se a técnica acertou e gerou o menor erro dentre as técnicas presentes na seleção.

Uma propagação de localização é considerada correta quando a distância entre a localização propagada e a localização real está dentro de um limiar de erro. Dessa forma, mesmo quando as propagações estão corretas, ainda se pode minimizar o erro.

Os valores das técnicas podem ser normalizados, considerando-se uma minimização do erro dentro do limiar estabelecido ou uma minimização do erro em geral. O processamento para encontrar os valores das colunas das técnicas é descrito a seguir:

- "Técnica 1", "Técnica 2", "Técnica 3" e "Técnica 4": recebem o valor normalizado por meio da função N3 ou da função N4, em que "min" é o parâmetro que guarda o menor erro encontrado dentre as quatro técnicas da seleção automática, para a fotografia em questão;

$$N3(x, \min) = \begin{cases} 1, & \text{Se } (x \leq 100) \text{ E } (x \leq \min) \\ 0, & \text{Caso contrário} \end{cases} \quad (4.27)$$

$$N4(x, \min) = \begin{cases} 1, & \text{Se } (x \leq 100) \text{ OU } (x \leq \min) \\ 0, & \text{Caso contrário} \end{cases} \quad (4.28)$$

Assim, ao final da segunda etapa, obtém-se uma grande tabela armazenando todos os metadados associados às fotografias e uma coluna para cada técnica de propagação, indicando se essa técnica deve ou não ser escolhida para a foto.

• Etapa 3 - Teste dos modelos lineares

Ao final da segunda etapa, verifica-se uma tabela com todos os metadados associados às fotografias que receberam propagação na etapa 1. Para testar a seleção automática das técnicas a partir de regressão, foi usada a validação cruzada. A tabela construída na etapa 2 é dividida em conjunto de treinamento e conjunto de teste para a regressão. O conjunto de treinamento servirá para a construção dos modelos lineares e será composto de 70% das linhas da tabela de metadados. A escolha das linhas que farão parte do conjunto de treinamento é feita aleatoriamente. Durante a construção do conjunto de treinamento, um *ranking* das técnicas é construído para ser usado no caso de empate entre técnicas a serem escolhidas.

Com a tabela dividida em dados de treinamento e dados de teste, os modelos lineares dos usuários são criados com base nos dados de treinamento. Com os modelos lineares criados, os metadados do conjunto de teste são usados para testá-los. Cada usuário tem um modelo linear representando cada técnica da seleção automática.

Assim, para cada linha do conjunto de teste é analisado o valor do modelo linear de cada técnica. O modelo linear que retornar o valor maior representará a técnica que deverá ser usada na propagação de localização para aquela linha do conjunto de teste.

Se a técnica escolhida pela seleção automática tem o valor 1 na tabela construída na etapa 2, considera-se um acerto da técnica de seleção automática com regressão linear múltipla. Em outras palavras, seja T1 a tabela com os valores dos modelos lineares construídos na regressão linear múltipla e T2 a tabela com os valores das técnicas na tabela construída na etapa 2. Então:

- $L1_n$ uma lista que representa a n-ésima linha ($L1_n \in T1$) no formato $[v1, v2, v3, v4]$;
- $L2_n$ uma lista que representa a n-ésima linha ($L2_n \in T2$) no formato $[t1, t2, t3, t4]$;
- $\text{indiceDoMaior}([x,y,z,w])$ a função que retorna a posição do maior valor encontrado, considerando o *ranking* das técnicas;
- $\text{linhaCorreta}(L1_i, L2_j)$ a função representada pela equação 4.29 a seguir:

$$\text{linhaCorreta}(L1_i, L2_j) = L2_j[\text{indiceDoMaior}(L1_i)] \quad (4.29)$$

Com base na função “linhaCorreta” e considerando-se a técnica proposta neste artigo, o acerto é definido pela equação 4.30:

$$\text{acerto} = \frac{\sum_{k=0}^n \text{linhaCorreta}(L1_k, L2_k)}{n} \quad (4.30)$$

4.2.3 Resultados

Como se mostrou nas seções anteriores, as técnicas de propagação de localização possuem comportamentos diferentes de acordo com o perfil do usuário, o que evidencia que são usadas diferentes técnicas em cada cenário de teste. Isso faz com que a técnica proposta neste trabalho possa ser aplicável em qualquer cenário de teste.

Então, para verificar a aplicabilidade da técnica proposta, foram testadas as seguintes hipóteses:

- H1: A técnica proposta apresenta uma porcentagem de minimização maior que a técnica com melhor resultado;
- H2: A técnica proposta apresenta uma porcentagem de minimização igual a técnica com melhor resultado;

A respeito da normalidade dos dados, foram feitos testes e foi utilizado o teste Student-t para os dados normais e o teste Wilcox (BOSLAUGH e WATTERS, 2008) para os demais. O teste de Shapiro e Wilk (1965) e o teste de Anderson e Darling (1952) foram usados para testar a normalidade.

Os resultados dos testes, após sua aplicação nos três cenários e usando as funções N3 e N4, definidas na etapa 2, estão configurados na tabela 4.2. Nota-se que a seleção automática de técnicas de propagação de localização melhora os resultados em cinco das seis variações de testes feitas. O único cenário em que os resultados foram ligeiramente inferiores foi com a função de normalização N3, que foca na minimização do erro de propagação dentro do limiar estabelecido. Considerando-se a normalização da função N4, que está mais direcionada com a minimização do erro de propagação e não necessariamente com a maximização do acerto, a seleção automática melhorou a cobertura em todos os cenários de teste.

Tabela 4.2. Confirmação das hipóteses

	Com a função N3	Com a função N4
Cenário 1	H1 é rejeitada H2 é rejeitada	H1 é aceita H2 é rejeitada
Cenário 2	H1 é aceita H2 é rejeitada	H1 é aceita H2 é rejeitada
Cenário 3	H1 é aceita H2 é rejeitada	H1 é aceita H2 é rejeitada

Os valores das diferenças médias entre a técnica proposta e a melhor técnica de cada cenário podem ser vistos na tabela 4.3. Nos outros cenários, houve aceitação da hipótese, ou seja, a técnica proposta apresentou uma porcentagem maior que a técnica com melhor resultado no cenário.

Tabela 4.3. Diferença média entre a seleção automática proposta e a melhor técnica do cenário

	Com a função N3	Com a função N4
Cenário 1	-0.56 %	0.23 %
Cenário 2	1.78 %	1.37 %
Cenário 3	1.66 %	3.13 %

Sendo assim, a técnica proposta neste trabalho apresenta bons resultados em todos os cenários testados, diferentemente das demais técnicas, pois, para cada cenário, existe uma abordagem que apresenta melhores resultados. Comprova-se, portanto, que, apesar de a seleção automática proposta não ter sido melhor em um cenário, na maioria dos casos ela apresentou melhorias com relação à técnica com os melhores resultados para o cenário específico testado.

4.3 Considerações finais

Neste capítulo, foram propostos dois algoritmos de propagação de localização geográfica em coleções de fotografias, foi feita uma análise comparativa com dois algoritmos existentes na literatura e foi proposta uma seleção automática de técnicas de propagação usando regressão linear múltipla.

Analisando-se os testes completos, pode-se afirmar que a escolha do algoritmo de propagação de localização adequado dependerá da métrica e do cenário de teste que melhor representem a realidade do sistema em questão. Além disso, a escolha do algoritmo também deve levar em consideração o perfil de cada usuário do sistema individualmente, ou seja, deve ser feita uma escolha de algoritmo personalizada para cada um.

A seleção automática é feita considerando-se o perfil de cada usuário e o contexto da fotografia. Com os testes realizados durante toda a análise, pode-se afirmar que a aplicação de regressão linear múltipla para selecionar automaticamente as técnicas analisadas melhorou a propagação de localização em cinco dos seis cenários de teste e fez com que ela se adaptasse a todos eles.

Além disso, dependendo da base de dados usada no processo, a mudança automática de técnica é importante, pois pode prevenir erros futuros com os modelos lineares por usuário. Isso faz com que a seleção automática apresentada neste trabalho seja autoadaptável ao perfil do usuário e, conseqüentemente, ao perfil da base de dados.

Capítulo 5 – Anotação de eventos

A organização de fotografias hodierna está aperfeiçoada e existem algumas abordagens que usam de artifícios, como as pessoas presentes nas fotografias, para ajudar a organização e recuperação das imagens. Alguns pesquisadores da área constataram que uma das formas utilizadas para organizar fotografias está no conceito de evento (LIM et al., 2003). Assim, automatizar a organização de fotografias em eventos significa diminuir o trabalho de uma tarefa que vem sendo realizada pelas pessoas.

As pesquisas na área seguem abordagens diferentes na organização de fotografias em eventos. Alguns trabalhos procuram a anotação do evento na fotografia, outros buscam o agrupamento de fotografias do mesmo evento, independentemente da anotação dos eventos. Essas abordagens de agrupamento e anotação podem englobar três tipos de eventos: eventos pessoais, eventos sociais e eventos colaborativos (FIGUEIRÊDO et al., 2012; BRENNER e IZQUIERDO, 2012; ZIGKOLIS et al., 2012). Os eventos pessoais são os que se referem a um acontecimento pessoal, como “aniversário de papai”. Os eventos sociais são aqueles que representam um acontecimento na sociedade, como, por exemplo, um concerto musical. Os eventos colaborativos englobam os eventos pessoais e sociais, porém suas informações são fornecidas pela colaboração dos usuários do sistema. Essa colaboração se dá quando os usuários cadastram as informações dos eventos (nome, data, local) no sistema. Este trabalho centra esforços na anotação de eventos sociais previamente conhecidos, podendo vir de fonte colaborativa ou não.

Muitos aparelhos de captura de imagens atuais possuem processamento digital de imagens acoplado, gerando informações cada vez mais claras e estruturadas, tais como os lugares das faces, se estão sorrindo ou não. Essas informações podem ser usadas para facilitar a anotação de pessoas e localização, por exemplo. Desse modo, as técnicas de anotação devem estar focadas no uso das informações associadas para encontrar a anotação desejada.

Além das informações geradas pelos dispositivos e pelas técnicas de anotação

existentes, os usuários também estão gerando conteúdos semiestruturados por meio de *tags* textuais associadas às suas fotografias. Dessa forma, além dos metadados associados às fotografias, existe também a informação gerada pelo usuário que possui a fotografia.

Para atenuar a tarefa de anotação de eventos em coleções de fotografias digitais, propõem-se um algoritmo de agrupamento de fotografias denominado Pegado e um algoritmo de propagação automática de anotação de eventos denominado Muído. O Pegado é baseado na localização, data e hora das fotografias. O Muído usa os grupos criados no Pegado e as *tags* anotadas nas fotografias para realizar a propagação automática de anotação de eventos.

Visando à necessidade de diminuição no trabalho de anotação de eventos em coleções de fotografias, as principais contribuições apresentadas neste capítulo são:

1. Pegado - um algoritmo de agrupamento de fotografias;
2. Análise comparativa entre as configurações possíveis de agrupamento temporal e espacial;
3. Muído - um algoritmo de propagação automática de anotação de eventos;
4. Análise de precisão e cobertura da propagação automática de eventos.

5.1 Agrupamento de fotografias

A organização de fotografias em eventos pode envolver dois problemas: a identificação do evento ao qual a foto pertence e o agrupamento de fotografias de um mesmo evento. O agrupamento é direcionado à recuperação de fotografias, pois não realiza nenhuma anotação. Dessa forma, uma técnica de agrupamento pode ser avaliada por meio da quantidade de grupos criados e da quantidade de fotografias de eventos distintos presentes em um mesmo grupo. O foco deste trabalho é a anotação de eventos e não o agrupamento de fotografias, logo o agrupamento foi direcionado para minimizar a quantidade de eventos distintos presentes em um mesmo grupo.

O algoritmo Pegado não faz uso de conjunto de treinamento. O modelo

aplicado pelo Pegado considera as informações espaciais e temporais para criar grupos com o algoritmo de agrupamento DBScan (ESTER et al., 1996). A informação espacial é necessária para que o Pegado realize o agrupamento, ou seja, caso uma fotografia não possua localização, ela não fará parte de nenhum grupo final criado pelo algoritmo.

O DBScan considera a distância entre os objetos para realizar o agrupamento. Quando os objetos possuem mais de um atributo a ser usado, o agrupamento pode ser feito considerando-se todos os atributos de uma única vez ou para cada atributo separadamente, criando subgrupos.

Considerando-se o contexto de fotografias e as dimensões espacial e temporal, o agrupamento pode ser feito das seguintes maneiras:

1. Com as duas dimensões;
2. Com a dimensão temporal inicialmente e, para cada grupo da dimensão temporal, um agrupamento espacial diferente;
3. Com a dimensão espacial inicialmente e, para cada grupo da dimensão espacial, um agrupamento temporal diferente.

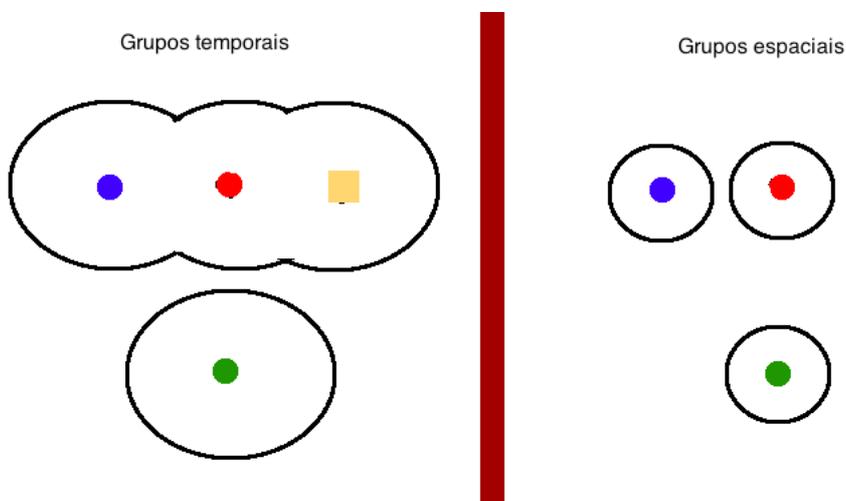
Como a informação temporal é algo consolidado nas fotografias, é feito o agrupamento da dimensão temporal e, depois, da dimensão espacial separadamente. Dessa forma, mesmo o Pegado tendo sido elaborado para fotografias com localização geográfica, as que não possuem localização ganham mais chances de estarem relacionadas com algum grupo temporal.

Na figura 5.1, evidencia-se um exemplo que mostra a vantagem do agrupamento temporal antes do espacial. Considere-se que a fotografia representada por um quadrado laranja não tenha localização geográfica e que os contornos pretos representem os grupos. Se o agrupamento espacial for feito em conjunto com o temporal ou antes do temporal, o resultado serão os grupos apresentados do lado direito na figura 5.1. Dessa forma, caso a fotografia faça parte do mesmo evento da representada pela bola azul, não será possível realizar a associação.

Considerando-se a abordagem de criar os grupos temporais e, depois, criar os grupos espaciais separadamente para cada grupo temporal, mesmo o resultado dos grupos finais sendo o mesmo do agrupamento em conjunto, existirá o grupo temporal

para fazer a ligação.

Figura 5.1. Grupos temporais e grupos espaciais



O algoritmo 5.1 representa o fluxo de execução do Pegado. Primeiramente, os grupos temporais são criados e, em seguida, os grupos espaciais, considerando-se cada grupo temporal existente.

Algoritmo 5.1. Validação do agrupamento de fotos

```
gruposTemporais = criarGruposTemporais(todasAsFotografias);  
salvarGrupos(gruposTemporais);  
for(grupo in gruposTemporais){  
    gruposTempoEspaciais = criarGruposEspaciais(grupo);  
    salvarGrupos(gruposTempoEspaciais);  
}
```

O agrupamento temporal ocorre como descrito na seção 4.1.2, que descreve o funcionamento da técnica de Agrupamento Temporal reproduzida para a validação das técnicas de propagação. O agrupamento espacial é feito de maneira similar ao temporal, porém a distância entre as fotografias é verificada por meio da distância geodésica das coordenadas geográficas das fotografias.

Além da distância entre os objetos, o DBscan também considera o número de objetos dentro de cada possível grupo para salvar ou não o grupo. Como o agrupamento feito pelo Pegado é direcionado para a propagação de anotação de eventos, não existe um número mínimo de fotografias por grupo, portanto uma fotografia isolada é considerada um grupo.

5.2 Propagação de eventos

Conforme exposto na seção anterior, a organização de fotografias em eventos pode envolver problemas de identificação do evento e do agrupamento de fotografias. A identificação do evento é voltada para associar uma determinada fotografia a um evento específico. Dessa forma, uma técnica de detecção de eventos pode ser avaliada por meio da precisão, cobertura e medida F (do inglês, *F-Measure*) das detecções realizadas automaticamente. O foco deste trabalho é a anotação de eventos; logo, a detecção de eventos foi especificada para maximizar os valores de precisão, cobertura e medida F. Essas métricas serão detalhadas na seção de validação deste capítulo.

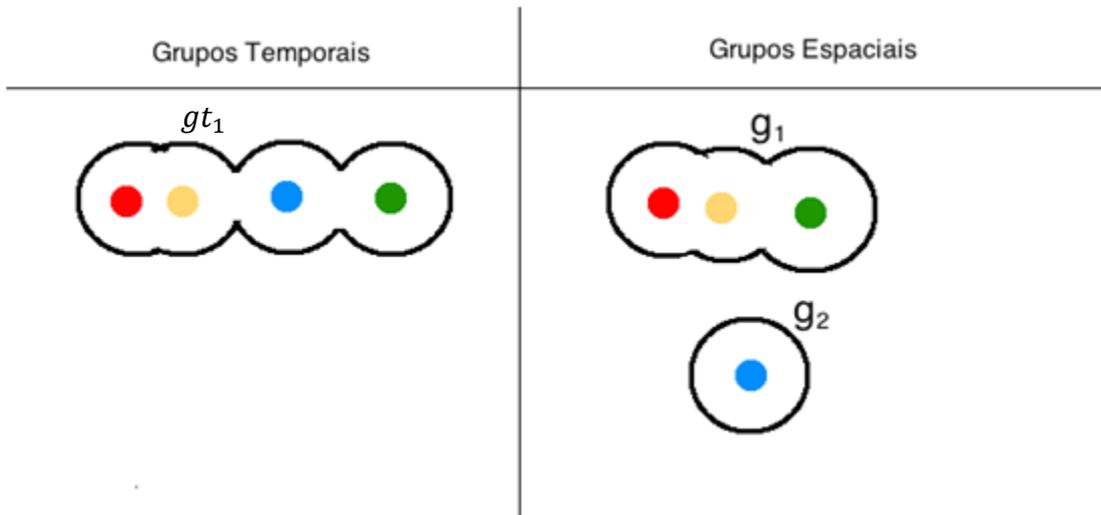
O Muído, algoritmo de anotação automática de eventos proposto neste trabalho, faz uso de um conjunto de treinamento composto por anotações de eventos. Essas anotações são usadas em conjunto com os grupos resultantes do modelo de agrupamento presente no Pegado. As anotações de eventos do conjunto de treinamento servirão para identificar o evento predominante nos grupos do Pegado. O evento predominante de um grupo é aquele que possui o maior número de fotografias dentro do grupo em questão.

Seja G o conjunto de grupos criado pelo algoritmo Pegado e g_i um grupo qualquer de G . O evento predominante de g_i será o evento eg_i (que possui o maior número de fotografias em g_i). Seja ft uma fotografia que não possui evento associado. O evento a ser anotado em ft é definido pela função *escolheEvento*, descrita a seguir:

$$escolheEvento(ft) = eg_i : ft \in g_i; \quad (5.1)$$

Na figura 5.2, é apresentado um exemplo de anotação de fotografia. Considerando-se a bola verde a fotografia ft sem evento associado, o Muído identificará o eg_i do grupo g_i em que ft está inserida. Dessa forma, o evento eg_i será associado a ft .

Figura 5.2. Anotação de evento em fotografia com localização



Assim como o Pegado, o algoritmo Muído é direcionado a fotografias com localização geográfica. Contudo, seja ft uma fotografia sem localização, GT o conjunto de grupos temporais, criados anteriormente aos grupos espaciais, gt_i um grupo temporal pertencente a GT e g_k^i um grupo espacial pertencente a G e criado a partir de gt_i . Se ft está inserida no intervalo temporal de algum gt_i e possui pelo menos uma tag em comum com as fotografias de gt_i , o algoritmo Muído encontra o grupo g_z^i relacionado com ft e anota o evento predominante eg_z^i em ft .

Figura 5.3. Anotação de eventos com TF-IDF

Tags		Grupos Temporais	Grupos Espaciais
●	RockInRio Metallica		
●	RockInRio LinkinPark		
●	Ivete Axé Maracanã		
●	RockInRio Scorpions		

Para encontrar o grupo g_z^i , o Muído calcula a semelhança da fotografia com todos os grupos g_k^i criados a partir de gt_i tal que $ft \in gt_i$. A semelhança é calculada por meio de uma função TF-IDF baseada em $tags$ textuais (RAJARAMAN e

ULLMAN, 2011). O TF é calculado por meio da frequência das *tags* em cada grupo g_k^i e o IDF das *tags* é calculado pela quantidade de grupos que as possuem. O cálculo é semelhante ao do TF-IDF tradicional, em que o termo é uma *tag* e o documento é um grupo criado pelo Pegado. Então, é escolhido o grupo g_z^i que apresentar o maior TF-IDF em relação à fotografia *ft*.

Na figura 5.3, é apresentado um exemplo do funcionamento da anotação de eventos em fotografias que não possuem localização, no qual a bola verde é a fotografia *ft* sem localização. Como *ft* tem pelo menos uma *tag* em comum com as fotografias de g_1^l , a anotação automática poderá acontecer. É calculado o TF-IDF dos grupos g_1^l e g_2^l , considerando-se as *tags* de *ft*. Como o grupo g_2^l não possui nenhuma *tag* em comum, o TF-IDF será zero. O grupo g_1^l tem *tag* em comum, logo o TF-IDF será maior que zero. Dessa forma, o g_z^l escolhido é o grupo g_1^l , e o evento anotado em *ft* será eg_1^l .

Figura 5.4. Fluxograma do algoritmo Muído



Na figura 5.4, é apresentada uma sumarização de como o algoritmo Muído funciona. Primeiramente, criam-se os grupos com o Pegado, depois se percorrem as fotografias com localização que não possuem evento associado e, para cada uma delas, anota-se o evento predominante do grupo. Em seguida, buscam-se as fotografias sem localização que possuem pelo menos uma *tag* em comum com o grupo temporal ao

qual pertence. Por último, para cada um das fotografias resultantes da busca anterior, anota-se o evento predominante do grupo a que pertence a maior semelhança (TF-IDF) com a fotografia.

5.3 Validação

Nesta seção, expõe-se o método utilizado para coletar os dados, as características sobre a base de fotografias usada para testar as técnicas, as métricas adotadas para analisar os resultados obtidos e, por fim, como as métricas foram validadas para gerar os resultados obtidos.

A validação da anotação de eventos foi feita individualmente, pois as técnicas encontradas na literatura não puderam ser replicadas de maneira fiel pela falta de informações de configuração das técnicas. Dessa forma, a análise da anotação de eventos proposta neste trabalho é feita com base nas métricas precisão, cobertura e medida F, detalhadas na subseção 5.3.3.

5.3.1 Base de dados

Neste trabalho, usou-se uma base de dados da literatura para extrair as fotografias usadas nos testes dos algoritmos propostos. Essa base de dados foi proposta por Reuter et al. (2014), extraída do Flickr e composta por fotografias que possuem a anotação dos seus respectivos eventos. Como o Pegado usa a localização para criar os grupos, apenas as fotografias dos eventos que possuem pelo menos uma fotografia com localização foram consideradas.

Dessa forma, a base filtrada resulta em 6.650 eventos e 131.551 fotografias. As fotografias presentes nessa base filtrada possuem *tags* textuais, data e hora da captura e localização geográfica na forma de latitude e longitude.

5.3.2 Validação do agrupamento de fotografias

A validação do agrupamento de fotografias com relação aos eventos foi supervisionada e usou como base a quantidade de fotografias pertencentes a eventos diferentes dentro de um mesmo grupo. Como o propósito do agrupamento é a propagação de eventos entre as fotografias do grupo, quanto menor a quantidade de fotografias de eventos diferentes, melhor.

Seja G o conjunto de grupos criado pelo algoritmo Pegado e g_i um grupo qualquer de G . O evento predominante de g_i será o evento que possui o maior número de fotografias em g_i . Seja $qtdFotosDif(x)$ a função que retorna a quantidade de fotografias distintas ao evento predominante de um grupo de fotografias x qualquer. A métrica usada para avaliar as configurações do agrupamento de Pegado é definida como:

$$PGr = \sum_{i=0}^{|G|} qtdFotosDif(g_i) \quad (5.2)$$

Para encontrar a melhor configuração de agrupamento, foram usados os seguintes valores para o agrupamento temporal: 30 minutos, 1 hora, 2 horas, 3 horas, 4 horas, 8 horas, 16 horas, 1 dia, 3 dias e 1 semana. Para o agrupamento espacial, foram considerados os seguintes valores: 20 metros, 50 metros, 100 metros, 500 metros, 1 quilômetro e 2 quilômetros. Os valores para os agrupamentos temporal e espacial foram escolhidos para abranger as configurações de duração temporal e espacial encontradas na literatura.

Algoritmo 5.2. Validação do agrupamento de fotos

```

valoresTemporais = {30min, 1h, 2h, 3h, 4h, 8h, 16h, 1d, 3d, 1sem};
valoresEspaciais = {20m, 50m, 100m, 500m, 1km, 2km};

for (replica = 0; replica < 15; replica++) {
    fotografiasAAgrupar = getFotografiasAAgrupar(30%);
    for(tempo in valoresTemporais){
        for(espaco in valoresEspaciais){
            grupos = Pegado(fotografiasAAgrupar, tempo, espaco);
            PGr = calculaPGr(grupos);
            salvarResultado(PGr, tempo, espaco);
            deletarGrupos(grupos);
        }
    }
}

```

O algoritmo 5.2 apresenta as etapas de coleta dos resultados. Para cada uma das 30 réplicas executadas, é coletada uma parcela aleatória de 30% das fotografias para realizar o agrupamento e verificar o valor de PGr. Após a parcela aleatória ter sido escolhida, o agrupamento é testado para cada combinação dos intervalos de tempo e das distâncias espaciais.

Ao fim da coleta, obtêm-se os valores de PGr para identificar qual a melhor configuração a ser usada no agrupamento espacial e temporal realizado por Pegado. Esses resultados são analisados com o teste estatístico adequado, dependendo da normalidade ou não dos dados, Teste-T ou Wilcoxon (BOSLAUGH e WATTERS, 2008), considerando-se um nível de significância $\alpha = 5\%$. Dessa forma, os testes garantem 95% de confiança nos resultados. Destarte, pode-se chegar a conclusões em relação à população e não somente para a amostra do experimento.

5.3.3 Validação da propagação de eventos

A validação da propagação automática de eventos foi supervisionada e usou como base algumas métricas conhecidas na literatura. As métricas adotadas foram precisão, cobertura e medida F (do inglês, *F-Measure*), em que cada uma delas é medida com base em três números:

1. S_c = Quantidade de propagações corretas;
2. S_e = Quantidade de propagações erradas;
3. S_n = Quantidade de fotografias que não receberam propagações.

As métricas são calculadas da seguinte forma:

- **Cobertura:**
$$\frac{S_c}{(S_c + S_e + S_n)}$$
 (5.4)

- **Precisão:**
$$\frac{S_c}{(S_c + S_e)}$$
 (5.5)

- **Medida F:**
$$2 \cdot \frac{\text{Precisão} \cdot \text{Cobertura}}{\text{Precisão} + \text{Cobertura}}$$
 (5.6)

Todas as métricas possuem importâncias individuais e tanto devem ser analisadas separadamente como em conjunto. A métrica nomeada cobertura indica o número de fotografias que receberam a localização dentro do limiar de erro adotado, considerando-se todo o conjunto de fotografias.

Em algumas fotografias, os algoritmos de propagação não conseguem encontrar localização. A métrica denominada precisão não leva em consideração essas fotografias que não receberam localização e aponta o quão correta é a propagação de um determinado algoritmo. A medida F é usada para analisar os valores de cobertura e

precisão em uma única métrica.

Como foi detalhado na seção 5.2, o algoritmo Muído precisa de um conjunto de treinamento composto pelas anotações dos eventos nas fotografias. Assim, a escolha das fotografias que fazem parte do conjunto de treinamento é crucial para que a representatividade dos dados seja mantida. Para garantir que o conjunto de treinamento seja imparcial, a escolha das fotografias que compõem esse conjunto é feita de maneira aleatória.

Os experimentos foram feitos por meio de validação cruzada, considerando-se diferentes números de *folds* para verificar a consistência do algoritmo. Cada configuração do número de *folds* possui trinta réplicas. O algoritmo 5.3 representa a dinâmica de como os experimentos de testes foram executados.

Algoritmo 5.3. Validação da propagação de eventos

```
Pegado(30min, 20m);
for (replica = 0; replica < 30; replica++) {
    fotografiasRemoverTemp = getFotografiasARemover();
    removerTemporariamente(fotografiasRemoverTemp);
    for(fotografia in fotografiasRemoverTemp){
        eventoSugerido = sugerirEvento(fotografia);
        salvarSugestao(eventoSugerido);
    }
    restaurarEventoOriginal(fotografiasRemoverTemp);
}
```

Primeiramente, o algoritmo Pegado cria os grupos das fotografias. Em seguida, para cada réplica, o conjunto de treinamento é escolhido aleatoriamente, e a anotação automática é feita para as demais fotografias. Como a base de dados possui os eventos de todas as fotografias, após todas as sugestões de eventos terem sido salvas, o evento sugerido é comparado ao evento real.

Uma propagação é considerada correta quando o evento sugerido é igual ao evento real. Para cada réplica, o conjunto de treinamento é definido pela função "getFotografiasARemover()", que retorna os identificadores das fotografias que não fazem parte do conjunto de treino.

Com os identificadores das fotografias do conjunto de teste, a anotação dos eventos dessas fotografias é removida temporariamente. Após a remoção da anotação

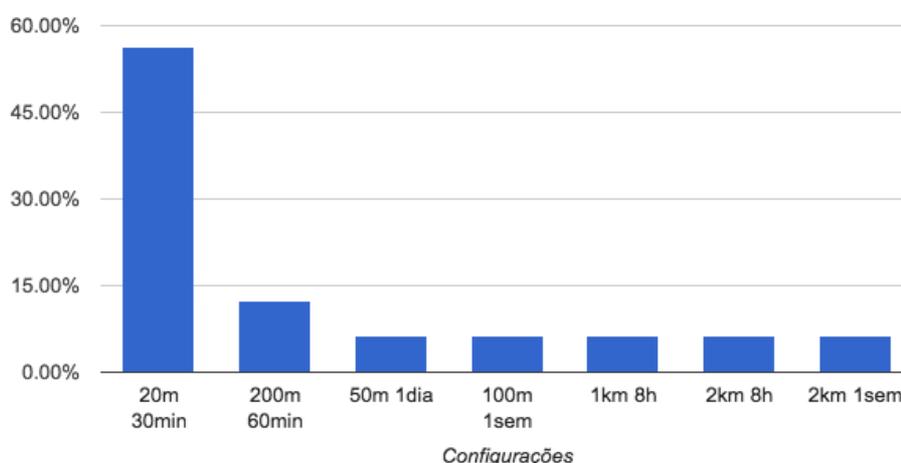
de eventos, o algoritmo Muído analisa todas as fotografias que não possuem eventos e faz a anotação do evento encontrado para cada uma. Após salvar todas as anotações realizadas, as anotações originais são restauradas para que a próxima réplica ocorra normalmente.

5.4 Resultados

Nesta seção, apresentam-se os resultados obtidos para os algoritmos Pegado e Muído em relação às métricas observadas e, então, discutem-se tais resultados. Os gráficos trazem os resultados obtidos pelos testes estatísticos executados. A maneira como as fotografias são agrupadas pelo Pegado influencia diretamente na propagação de localização realizada por Muído. Portanto, encontrar a configuração adequada para o agrupamento de fotografias de um mesmo evento é fundamental para que as propagações de anotação ocorram de maneira adequada.

Na figura 5.5, os resultados das configurações de agrupamento em Pegado são apresentados. O eixo vertical indica a porcentagem de réplicas em que aquela configuração obteve o melhor resultado. Dessa forma, o eixo horizontal representa as configurações de agrupamento.

Figura 5.5. Melhores configurações de agrupamento em Pegado

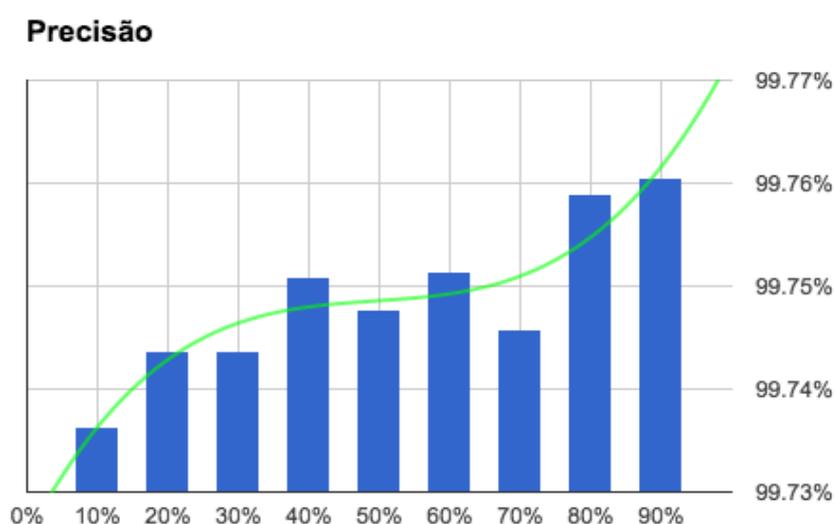


As configurações que não constam no gráfico não obtiveram o melhor resultado em nenhuma réplica. É notório que a configuração de agrupamento com 30

minutos de variação temporal e 20 metros de variação espacial é a mais indicada para agrupar fotografias de um mesmo evento, minimizando a quantidade de fotografias de eventos distintos no mesmo grupo.

Com o resultado dos testes para as configurações do algoritmo Pegado, os testes do algoritmo Muído puderam ser executados. Assim, as figuras 5.6, 5.7, 5.8, 5.9 e 5.10 apresentam os resultados do Muído com a configuração de agrupamento em Pegado de 30 minutos de variação temporal e 20 metros de variação espacial.

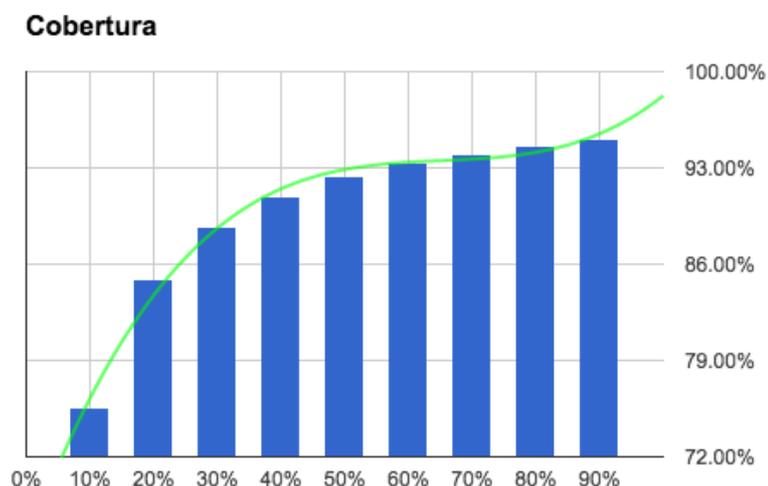
Figura 5.6. Precisão da anotação automática em fotografias com localização



Na figura 5.6, são apresentados os resultados referentes à precisão do algoritmo Muído na propagação de anotação de eventos em fotografias com localização geográfica. Esses resultados são relacionados com o conjunto de treinamento usado na propagação de anotação de eventos. Como a variação da precisão foi baixa, menos que 1%, ao diminuir o conjunto de treinamento até 10%, pode-se concluir que a precisão é estável e bastante elevada. A linha verde no gráfico aponta uma regressão polinomial dos resultados.

Na figura 5.7, são apresentados os resultados referentes à cobertura do algoritmo Muído na propagação de anotação de eventos em fotografias com localização geográfica. Assim como os resultados de precisão, eles estão relacionados com o conjunto de treinamento. É natural que, com a diminuição do conjunto de treinamento, a cobertura diminua, porém a cobertura só apresenta uma diminuição considerável com 10% de treinamento.

Figura 5.7. Cobertura da anotação automática em fotografias com localização



Apesar da diminuição considerável para 10% de treinamento, o valor alcançado ainda é alto. Dessa forma, o Muído se torna aplicável a coleções de fotografias com poucas anotações. A linha verde no gráfico também representa uma regressão polinomial dos resultados.

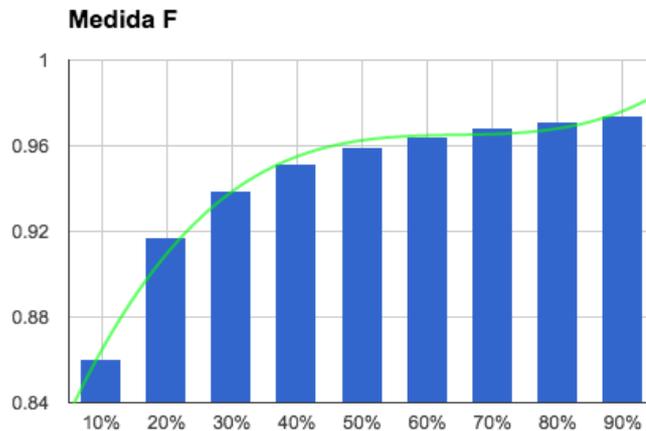
Na figura 5.8, são apresentados os resultados referentes à combinação da precisão e cobertura presentes nas figuras 5.6 e 5.7, respectivamente. Essa combinação na medida F tem como finalidade analisar as duas métricas em conjunto. Assim como os resultados de Precisão e Cobertura, os resultados da Medida F estão relacionados com o conjunto de treinamento.

Em razão dos altos valores de precisão e cobertura, naturalmente a Medida F também apresentou resultados altos. Mesmo com um conjunto de treinamento pequeno, de 10%, o resultado ainda foi próximo de 1. Assim como os demais gráficos, a linha verde indica a regressão polinomial dos resultados.

Para as métricas observadas, o Muído apresentou bons resultados na propagação de anotação de eventos em fotografias com localização geográfica. Os resultados foram constantes nas três métricas e considerados bons, mesmo com a diminuição do conjunto de treinamento. Considerando-se o contexto de fotografias sem localização geográfica, o Muído realiza a propagação por meio de semelhança de

tags.

Figura 5.8. Medida F da anotação automática em fotografias com localização



Como a propagação se dá em função de *tags*, a similaridade das fotografias com os grupos espaciais independe de conjunto de treinamento. Consequentemente, a propagação só apresentará um resultado diferente se o evento predominante do grupo espacial for diferente. Como a propagação de eventos para fotografias sem localização se processa após a propagação para fotografias com localização e os grupos espaciais quase sempre apresentam os mesmos eventos predominantes, os resultados da propagação de anotação de eventos em fotografias sem localização são constantes. Dessa forma, eles são representados pela média dos resultados de cada réplica.

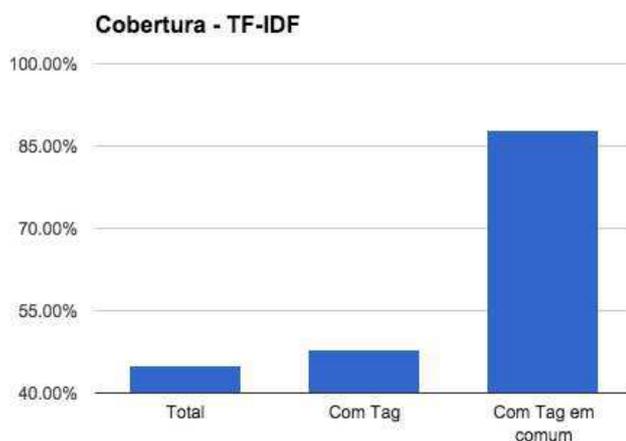
As métricas foram analisadas, considerando-se os seguintes conjuntos:

1. Total: todas as fotografias sem localização relacionadas com os eventos existentes na base de dados;
2. Com *tag*: subconjunto de Total, considerando-se as fotografias que possuem pelo menos uma *tag* em comum com alguma fotografia com localização;
3. Com *tag* em comum: subconjunto de Total, considerando-se as fotografias que possuem pelo menos uma *tag* em comum com pelo menos um grupo espacial criado a partir do grupo temporal a que a fotografia pertence.

Como a precisão leva em conta apenas as anotações automáticas feitas e não o conjunto inteiro, o valor médio foi o mesmo para os três cenários de teste analisados. O

valor médio da precisão foi de 88%. A cobertura, apresentada na figura 5.9, apresentou resultados bem distintos com a mudança do cenário de análise.

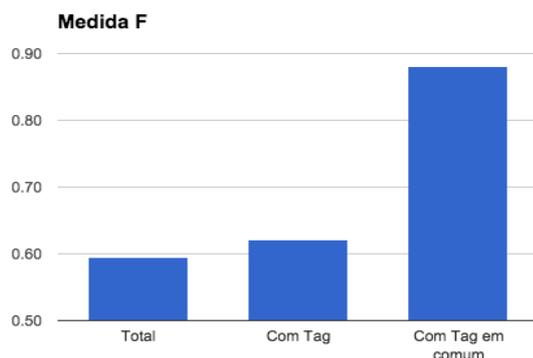
Figura 5.9. Cobertura da anotação automática em fotografias sem localização



Considerando-se o cenário com o conjunto Total, a cobertura apresentou um resultado de 44%. Para o conjunto Com *tag*, o resultado aumenta um pouco e fica em torno dos 49%. Considerando-se o conjunto Com *tag* em comum, o resultado aumenta consideravelmente e sobe para quase 90%. Dessa forma, fica claro que o conjunto de propagação de eventos por meio de *tags* só ocorre de maneira adequada quando a fotografia possui pelo menos uma *tag* em comum com os possíveis grupos espaciais.

A análise da Medida F, ilustrada na figura 5.10, apresenta ótimos resultados em relação ao conjunto Com *tag* em comum, porém não mostra melhorias significativas nos demais conjuntos. Dessa forma, o Muído pode ser usado para propagar anotação de eventos em fotografias sem localização. As anotações feitas terão boas chances de serem corretas (precisão de 88%), porém o alcance para fotografias sem localização não será tão alto, pois dependerá da existência de alguma *tag* em comum com os grupos espaciais provenientes do grupo temporal ao qual a fotografia pertence.

Figura 5.10. Medida F da anotação automática em fotografias sem localização



5.5 Considerações finais

Neste capítulo, apresentou-se uma abordagem para melhorar a anotação de eventos sociais em fotografias. O algoritmo Pegado, de agrupamento de fotografias, e o algoritmo Muído, de propagação de anotação de eventos, foram propostos e uma análise de cada um foi feita separadamente. A do Pegado foi feita em relação à quantidade de fotografias de eventos distintos em um mesmo grupo e a do Muído foi baseada nas métricas precisão, cobertura e medida F.

Com a análise do algoritmo Pegado, fica evidente que o agrupamento de fotografias deve ser feito considerando-se uma variação temporal de 30 minutos e 20 metros para minimizar a quantidade de fotografias de eventos distintos em um mesmo grupo. A análise do algoritmo Muído permite afirmar que a propagação de anotação de eventos fazendo uso do agrupamento feito pelo algoritmo Pegado é uma boa estratégia para fotografias com localização geográfica.

Com esses resultados, fica clara a importância da informação da localização geográfica nos sistemas de gerenciamento de fotografias. Para as fotografias que não possuem localização, a propagação de eventos para fotografias sem localização revela-se precisa, porém com baixa cobertura. Além disso, fica evidente a relação de eventos com as questões "Onde?" e "Quando?".

Capítulo 6 – Protótipo de ferramenta

Neste capítulo, a ferramenta PG++, desenvolvida durante a pesquisa, é detalhada. A PG++ centra esforços em melhorar a anotação de localização e eventos sociais em coleções de fotografias. Na ferramenta, a organização e recuperação das fotos toma como centro duas das questões identificadas por Naaman et al. (2004): "Onde?" e "Quando?". As duas questões servem para guiar as anotações realizadas nas fotografias por meio da ferramenta: a anotação de localização ("Onde?") e a anotação de eventos ("Onde?" e "Quando?"). Pelo fato de a anotação de eventos ocorrer em fotografias que possuem localização, a PG++ realiza primeiro a anotação de localização.

Em consequência da ausência de dispositivos GPS em algumas câmeras e da inconsistência do serviço em locais fechados, cabe ao usuário anotar a localização das fotografias manualmente. Para atenuar esse trabalho, técnicas de propagação de localização foram propostas, porém o funcionamento delas é dependente do conjunto de treinamento.

Considerando-se o cenário de usuários com câmeras fotográficas sem um dispositivo GPS, essa anotação se torna bastante trabalhosa, pois o conjunto de treinamento inicial para as técnicas de propagação existente será vazio. Nesse caso, a PG++ busca diminuir o trabalho do usuário por meio de uma abordagem diferente quanto ao problema de anotação de localização em fotos. Além de utilizar a técnica denominada Intuca para propagar localização, a ferramenta busca minimizar a quantidade de fotografias que precisam ser anotadas manualmente pelo usuário.

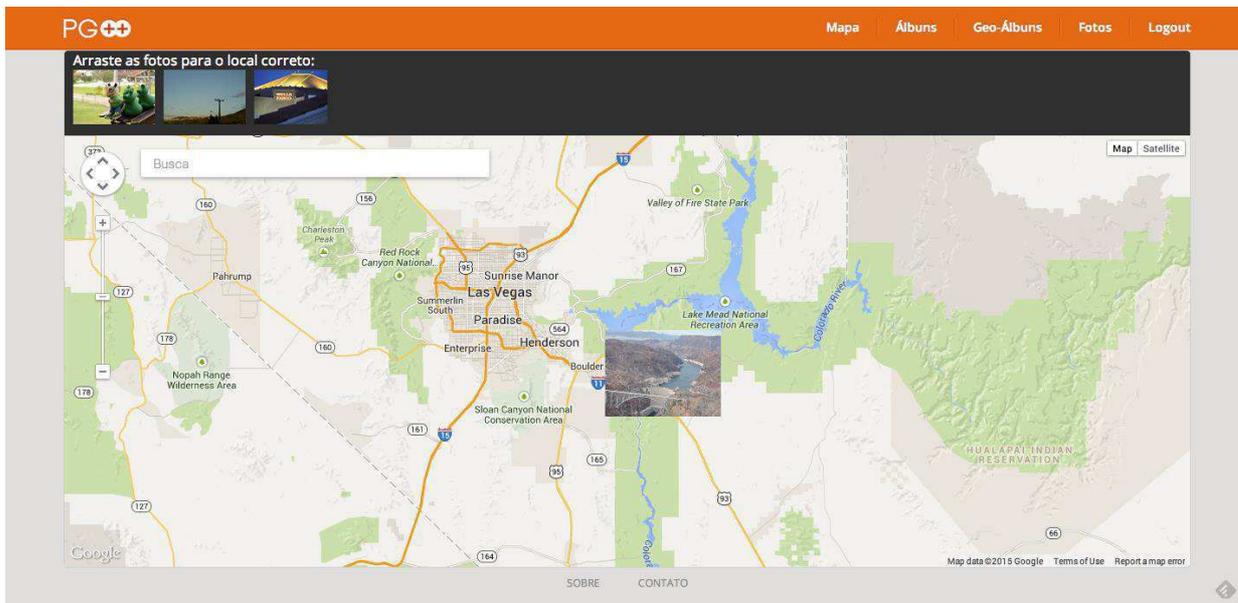
Após a propagação da localização e da anotação manual das fotografias, a PG++ utiliza a técnica Muído para encontrar o evento ao qual a fotografia pertence. Com as informações de localização e evento, a ferramenta cria alguns álbuns automaticamente para facilitar a recuperação de fotografias na coleção do usuário. A tela de *login* e cadastro é apresentada na figura 6.1. O usuário pode usar suas credenciais para acessar sua coleção de fotografias ou criar novas credenciais para ter

acesso completo à ferramenta. A tela inicial após o *login* pode ser visualizada na figura 6.2, onde o usuário poderá acessar sua coleção de fotografias por meio de um mapa.

Figura 6.1. Tela de *login* e cadastro da ferramenta



Figura 6.2. Tela inicial da ferramenta



6.1 Minimização da quantidade de fotografias a serem anotadas

Nesta seção, apresenta-se a técnica proposta para minimizar a quantidade de fotografias anotadas pelo usuário. O algoritmo adota a hipótese de que, tratando-se de propagação de localização, a probabilidade de encontrar localização para todas as fotografias é muito baixa. Considerando-se o contexto em que o usuário tem que anotar a localização de fotografias manualmente, o algoritmo de minimização se torna útil, mesmo nos casos em que o conjunto de treinamento não é vazio. Essa redução se verifica por meio da identificação de fotografias âncora para diminuir a quantidade de fotografias que precisam ser anotadas manualmente pelo usuário.

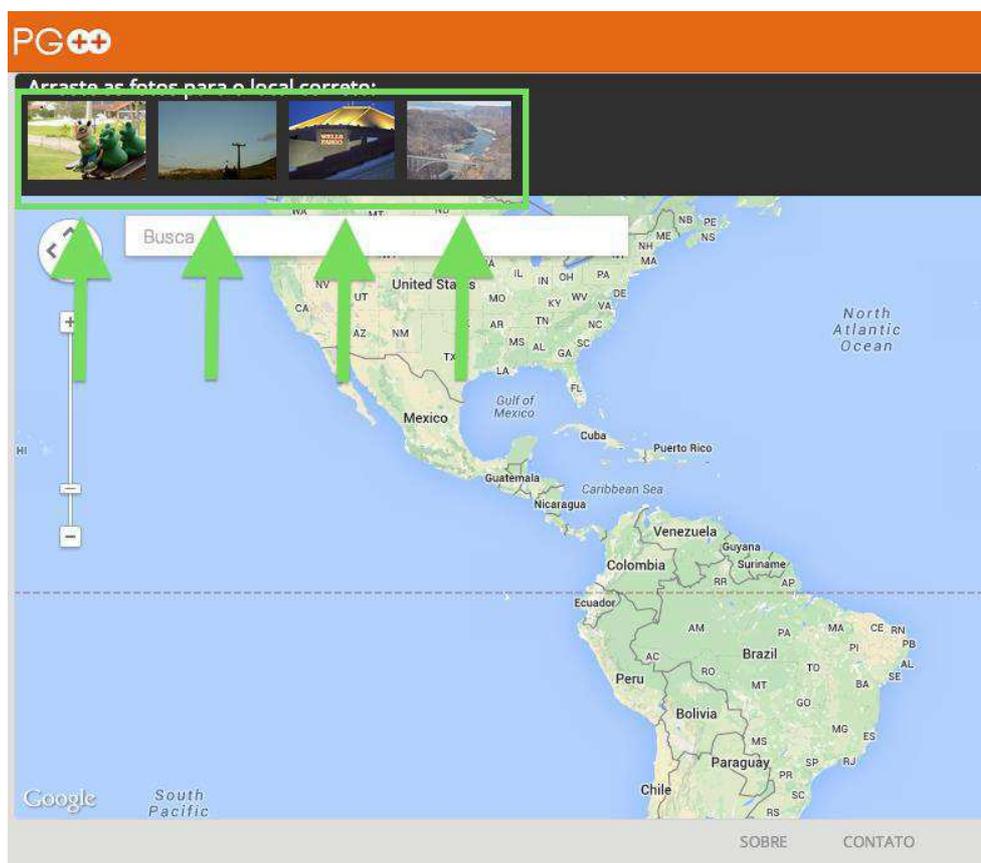
A técnica de minimização é acionada logo após a extração de metadados das fotos. Todas as fotos são agrupadas em grupos temporais, nos quais as fotos pertencentes a esses grupos são consecutivas. O algoritmo 6.1 apresenta como as fotos são divididas nos seus devidos grupos. Elas são ordenadas pela data de captura e separadas em grupos com base na distância temporal entre fotografias consecutivas. A divisão se processa de tal forma que os grupos possuam fotos consecutivas com distância temporal menor que o tempo máximo t_{max} estabelecido. O valor dessa variável foi estudado por Lacerda et al. (2013) e, com base nesse estudo, o valor definido para a ferramenta é de 60 minutos.

Algoritmo 6.1. Agrupamento das fotografias para minimização da anotação manual

```
fotosOrdenadas = conjunto de fotos ordenadas pela data em que foram tiradas;
dataDaUltimaFoto = null;
clusterAtual = new array;
clusters = new array;
while(não chegar ao fim de fotosOrdenadas) {
    p = leia foto atual;
    if(dataDaUltimaFoto = null OR (p.dataDeCaptura - dataDaUltimaFoto) < tmax) {
        clusterAtual.adicionar(p);
        dataDaUltimaFoto = p.dataDeCaptura;
    } else {
        clusters.adicionar(clusterAtual);
        currentCluster = new array;
        dataDaUltimaFoto = null;
    }
}
return clusters;
```

Os grupos são separados para que o usuário faça a anotação da localização. A técnica usa a foto no centro do grupo temporal para representar o grupo. Ao anotar a localização de cada uma dessas fotos centrais, o usuário também estará anotando a localização das fotos nos respectivos grupos temporais. A anotação dos grupos é mostrada na figura 6.3, onde a seta verde indica o grupo de fotografias que precisa ser anotado. Para realizar a anotação, o usuário deve clicar no grupo desejado e arrastá-lo para a localização no mapa onde a fotografia foi capturada.

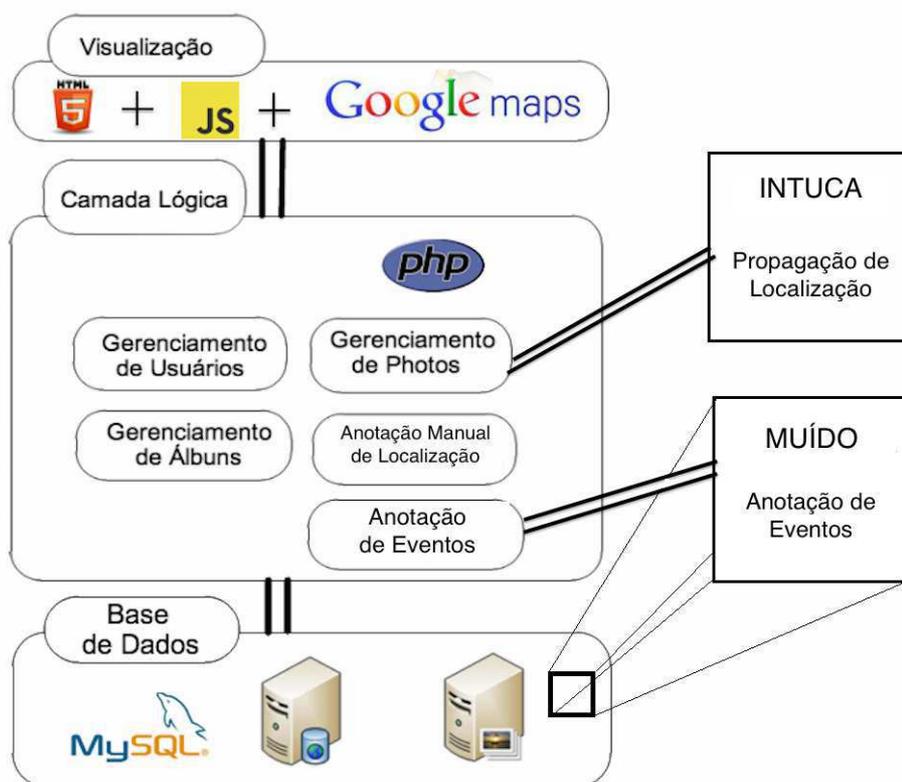
Figura 6.3. Tela de anotação de localização dos grupos



6.2 A arquitetura da ferramenta PG++

Na figura 6.4, a arquitetura da ferramenta é exposta. A camada de visualização é composta por HTML 5, Javascript e GoogleMaps. A combinação dessas três tecnologias permite algumas facilidades visuais que auxiliam na anotação da localização. Na camada lógica, os módulos de gerenciamento de usuários, fotos e álbuns realizam as atividades de cadastro, leitura, atualização e remoção em bases de dados relacionais.

Figura 6.4. Arquitetura da ferramenta PG++



A comunicação entre a camada de visualização e a camada lógica ocorre via requisições HTTP, em que a resposta é codificada em JSON. O módulo de anotação de localização é o responsável por aplicar a técnica Intuca e, em seguida, a técnica de minimização. O módulo de anotação de eventos é responsável por gerenciar as anotações de eventos das fotografias e requisitar a execução da técnica Muído.

A técnica Intuca realiza a comunicação por meio de requisições HTTP. Essas requisições partem do módulo de gerenciamento de fotos para Intuca, que funciona como um módulo externo. Esse módulo externo responde à requisição com a localização das fotografias, e o módulo de gerenciamento envia para a camada de armazenamento realizar as operações relativas ao arquivamento dos dados.

A técnica Muído é implementada diretamente na camada de armazenamento. A comunicação com o módulo de anotação de eventos é feita por meio do módulo de

armazenamento. O módulo de anotação de eventos requisita a execução, e o módulo de armazenamento executa os comandos SQL necessários para realizar as anotações.

A camada de armazenamento é a junção do banco de dados relacional, no qual os metadados são armazenados, com o sistema de arquivos, no qual as fotos são armazenadas. A camada lógica envia as informações para a camada de armazenamento realizar as operações necessárias ao arquivamento das fotos.

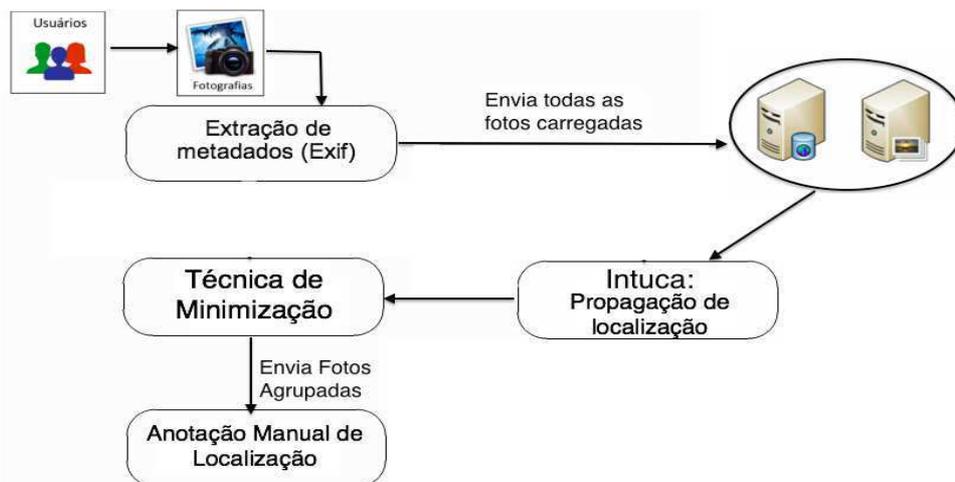
6.3 Anotação de localização

A ferramenta armazena as fotos no sistema de arquivos e todos os metadados em um banco de dados objeto-relacional com suporte a geoprocessamento. Após as informações relacionadas ao fato de as fotos terem sido extraídas e armazenadas, a técnica de minimização identifica as fotografias que não receberam a localização de possíveis fotos previamente armazenadas.

Essas fotografias identificadas são agrupadas e uma delas é passada para o usuário anotar a localização. Para que a anotação seja feita, ele deve identificar o local correto no mapa e, em seguida, arrastar a fotografia para lá. A localização anotada é passada para as demais fotografias, evitando-se, assim, que o usuário faça anotações repetidas ou que tenha o trabalho de agrupar manualmente as fotografias pertencentes a um mesmo local.

Na figura 6.5, é apresentada a ideia de minimização da quantidade de fotografias que precisam ser anotadas manualmente na PG++. O usuário carrega as fotos no sistema e, após a propagação de localização por meio da técnica Intuca, são identificadas quais fotos precisam ser anotadas manualmente para que todas recebam uma localização geográfica. Esse cálculo é feito com os metadados relacionados às fotografias.

Figura 6.5. Minimização da quantidade de fotografias anotadas manualmente



6.3.1 Extração de informações das fotos

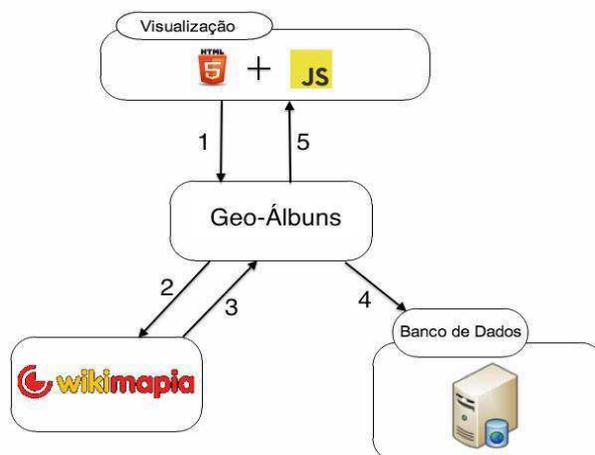
Para que o processamento das localizações das fotografias seja feito corretamente, a ferramenta precisa interpretar os dados embutidos nos arquivos. Dessa forma, é considerada a especificação “*Exchangeable image file format*” (Exif), seguida pela maioria dos fabricantes de câmeras digitais. Essa especificação dita a maneira como os metadados são etiquetados nas fotos e está presente nos formatos de imagem JPEG e TIFF. Por meio dessa especificação, a ferramenta concentra-se em extrair informações sobre o "Onde?".

Caso a foto não possua informação alguma seguindo a especificação Exif, a ferramenta armazena a data em que a foto foi salva no sistema como referência para associar à pergunta "Quando?".

6.3.2 Organização automática de álbuns

Para que a organização dessas fotos também seja melhorada e feita considerando-se a localização, a ferramenta cria automaticamente os chamados geoálbuns. Esses álbuns de fotos são criados com base nas localizações, e as fotos são agrupadas de acordo com suas respectivas regiões. Assim que a ferramenta recebe uma nova localização para ser armazenada, essas regiões são detectadas e associadas a cada foto.

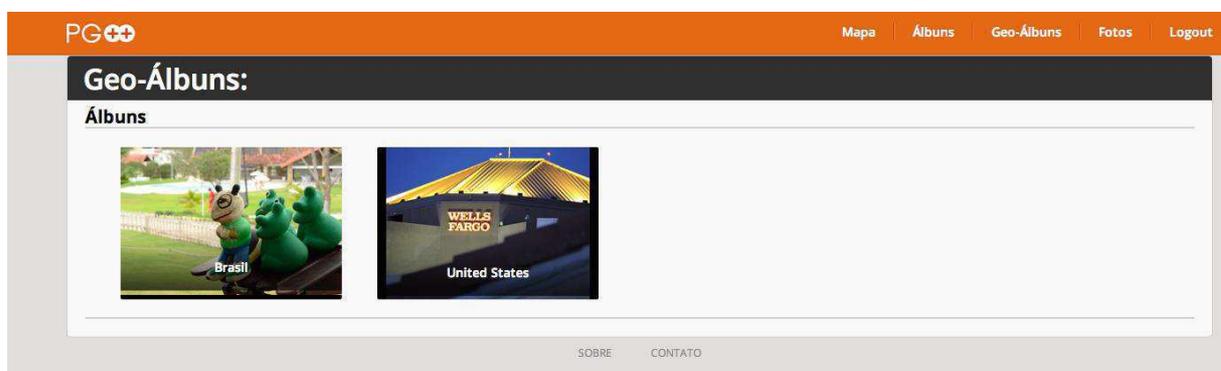
Figura 6.6. Criação dos geoálbuns



A detecção das regiões ocorre de maneira automática por meio do serviço prestado pelo Wikimapia. Uma requisição é feita ao Wikimapia para encontrar o local mais próximo à coordenada geográfica passada como parâmetro. A ferramenta interpreta o arquivo recebido e extrai as regiões políticas relacionadas com aquele local (país, estado, cidade). Essas regiões são armazenadas e mapeadas com o identificador único oriundo do Wikimapia para evitar duplicatas. Cada região tem um nível de profundidade associado e esses níveis são usados para definir a hierarquia de regiões. Uma região é filha de outra quando tem interseção com outra de nível menor.

Com as regiões salvas, a apresentação ao usuário é feita de maneira hierárquica. Na figura 6.6, é apresentada a sequência de passos para a criação dos geoálbuns. Na figura 6.7, é mostrada a interface de apresentação dos geoálbuns criados.

Figura 6.7. Interface de apresentação dos geoálbuns



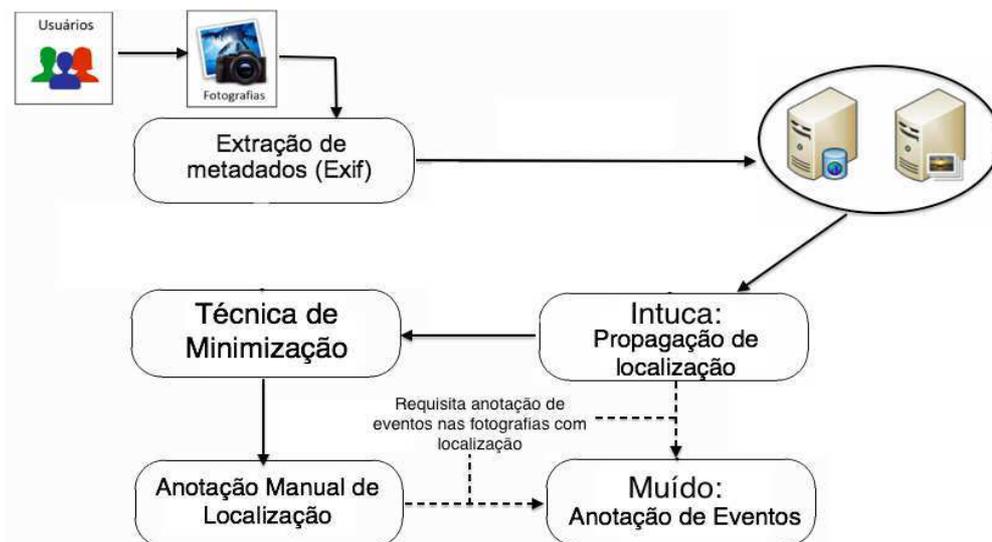
6.4 Anotação de eventos

Em decorrência de a técnica Muído ter sido proposta com a premissa de que as fotografias possuem localização geográfica, a anotação de eventos se dá após o armazenamento dessas localizações no banco de dados. Com as informações de localização, data e hora das fotografias armazenadas no banco de dados, a anotação de eventos é acionada.

As fotografias são agrupadas pela técnica Muído. Caso algum grupo de fotografias não possua nenhum evento relacionado, ele será destacado para que o usuário possa fazer a anotação manual do evento. Ao anotar o evento de alguma fotografia presente nos grupos, a técnica Muído é processada para atualizar as anotações automáticas feitas.

Na figura 6.8, é apresentado o momento em que a anotação de eventos é requisitada para as fotografias com localização. O usuário carrega as fotos no sistema e, após a propagação de localização por meio da Intuca, são identificadas quais fotos precisam receber o evento. Com essa identificação, a técnica Muído é acionada para encontrar o evento dessas fotografias.

Figura 6.8. Anotação de Eventos



6.4 Considerações finais

Neste capítulo, foi apresentada a PG++, uma ferramenta destinada a auxiliar a anotação de localização e eventos em coleções de fotografias. Ao contrário das técnicas e ferramentas existentes, a PG++ não só aplica técnicas de propagação de localização, como também utiliza uma técnica para minimizar a quantidade de fotografias anotadas manualmente. Tal abordagem de minimização favorece coleções com poucas fotografias dispondo de anotação de localização. A PG++ também organiza as fotografias automaticamente em geoálbuns para facilitar a recuperação delas pelo usuário.

Além da anotação de localização, a ferramenta auxilia na anotação de eventos nas fotografias do usuário. A PG++ cria álbuns dinâmicos baseados nos eventos das fotografias para aperfeiçoar a recuperação delas. Constata-se, portanto, que se trata de uma ferramenta completa que engloba desde a propagação e anotação de localização, questão "Onde?", até a anotação de eventos, questões "Onde?" e "Quando?".

Capítulo 7 – Conclusões e trabalhos futuros

A popularização dos dispositivos com câmeras fotográficas, em conjunto com a diminuição do custo de armazenamento, fez com que o número de fotografias aumentasse rapidamente. Esse grande número de fotografias dificulta a tarefa de anotar e catalogar tais arquivos, fazendo necessária a presença de mecanismos que diminuam o trabalho do usuário no gerenciamento desse material. Neste trabalho, foi proposto um ambiente para gerenciamento de fotografias com foco na localização e no evento das fotografias.

Para o gerenciamento com base na localização, dois algoritmos de propagação de localização geográfica em coleções de fotografias foram propostos: Trajetória e Eventos Compartilhados. O primeiro encontra a possível trajetória do usuário e, por meio de interpolação linear, infere a posição de uma fotografia. O segundo processa a localização das fotografias por meio de eventos pessoais e compartilhados por mais de um usuário, considerando-se que fotografias pertencentes a um mesmo evento também pertencem à mesma localização. A validação dos dois algoritmos se deu em um estudo comparativo com duas técnicas de propagação existentes na literatura (LACERDA et al., 2013; GONG et al., 2011).

Com os resultados empíricos obtidos, pode-se afirmar que a escolha do algoritmo de propagação de localização adequado dependerá da métrica e do cenário de teste que melhor representem a realidade do sistema em questão. Além disso, a escolha do algoritmo também deve levar em consideração o perfil de cada usuário do sistema individualmente, ou seja, deve ser feita uma escolha de algoritmo personalizada para cada usuário.

Ainda em relação à localização, foi proposta uma seleção automática de técnicas de propagação usando regressão linear múltipla. A seleção é feita considerando-se o perfil de cada usuário e o contexto em que a fotografia está inserida. Uma tabela de metadados das fotografias é construída e cada técnica de propagação de localização possui uma regressão linear múltipla, indicando se a técnica deve ser usada ou não. Dessa forma, cada usuário armazena uma regressão linear múltipla para cada

técnica a fim de que a seleção ocorra conforme seu perfil. Os resultados dos testes realizados durante toda a análise permitem afirmar que a aplicação de regressão linear múltipla para selecionar automaticamente as técnicas analisadas melhorou a propagação de localização em cinco dos seis cenários de teste e fez com que ela se adaptasse a todos eles. Essa mudança automática de técnica é importante, pois pode prevenir erros futuros com os modelos lineares por usuário. Isso faz com que a seleção automática apresentada neste trabalho seja autoadaptável ao perfil do usuário e, conseqüentemente, com o perfil da base de dados.

Para a anotação de eventos sociais, os algoritmos Pegado e Muído foram propostos. O Pegado realiza o agrupamento de fotografias baseando-se na similaridade espacial e temporal. O agrupamento é feito com a técnica DBScan. Primeiramente, é feito o agrupamento temporal e, em seguida, o agrupamento espacial. Com a análise do algoritmo Pegado, ficou evidente que o agrupamento de fotografias deve ser feito considerando-se uma variação temporal de 30 minutos e 20 metros para minimizar a quantidade de fotografias de eventos distintos em um mesmo grupo.

O algoritmo Muído realiza a propagação de anotação de eventos por meio dos grupos de fotografias criados pelo Pegado. Os dois algoritmos foram projetados para fotografias que possuem localização geográfica, porém, caso não a possuam, o Muído realiza uma similaridade por meio das *tags* das fotografias, usando a similaridade de cosseno em conjunto com a indexação *tf-idf*. A análise do algoritmo Muído permite afirmar que a propagação de anotação de eventos fazendo uso do agrupamento feito pelo algoritmo Pegado é uma boa estratégia para fotografias com localização geográfica.

Cabe ressaltar, com base nos resultados obtidos, a importância da informação da localização geográfica nos sistemas de gerenciamento de fotografias. Para as que não possuem localização, a propagação de eventos mostrou-se precisa, porém com baixa cobertura. Além disso, ficou evidente a relação de eventos com as questões "Onde?" e "Quando?".

Neste trabalho, também foi desenvolvida a ferramenta PG++, um protótipo que visa aplicar as técnicas de anotação de localização e eventos sociais propostas e

também procura minimizar a quantidade de fotografias que precisam ser anotadas pelo usuário.

Diferentemente das técnicas de propagação existentes, a técnica de minimização proposta na PG++ auxilia na anotação de fotografias, considerando-se uma coleção com poucas fotografias dispondo de anotação de localização. A PG++ também usa a localização das fotografias para criar geoálbuns automaticamente e facilitar a recuperação delas pelo usuário.

Além da ajuda na anotação de localização, a ferramenta auxilia na anotação de eventos na coleção de fotografias do usuário. Com a anotação automática dos eventos, a coleção é organizada em álbuns dinâmicos que facilitam a recuperação das fotografias.

7.1 Contribuições

As principais contribuições desta pesquisa são:

1. Duas técnicas de propagação de localização;
 - 1.1. Análise comparativa das técnicas propostas com duas técnicas da literatura;
2. Uma seleção automática de técnicas de propagação de localização com base no perfil do usuário e/ou da base de dados;
3. Uma técnica de agrupamento espaço-temporal de fotografias;
4. Uma técnica de detecção de eventos sociais;
5. PG++: um protótipo de ferramenta *web* para gerenciamento de fotografias;
 - 5.1. Agrupamento de fotografias para minimizar a quantidade de fotografias que necessitam de anotação manual;
 - 5.2. Organização automática em álbuns geográficos.

7.2 Trabalhos futuros

Por meio dos resultados obtidos nesta pesquisa, observaram-se os seguintes trabalhos futuros:

1. Ampliar as técnicas envolvendo localização para trabalharem também com localização descrita pelo usuário, além de coordenadas de latitude e longitude;
2. Aprimorar os parâmetros de agrupamento em Pegado para gerar um número de grupos mais semelhante ao número de eventos;
 - 2.1. Explorar a parametrização individualizada por usuário, ou por região, ou por período de tempo;
3. Propor uma técnica para identificar o nome de eventos desconhecidos por meio dos grupos formados pelo algoritmo Pegado;
4. Aprimorar a função de similaridade das fotografias com os grupos por meio de pré-processamento de *tags*;
5. Ampliar a ferramenta para suportar a exportação de fotografia com as anotações, seguindo um padrão existente, por exemplo, EXIF;
6. Analisar o desempenho de processamento das técnicas propostas;
7. Inserir as técnicas propostas em um cenário distribuído com tecnologias como o *Hadoop*.

Referências bibliográficas

ALTMAN, N. S. **An Introduction to Kernel and Nearest-neighbor Nonparametric Regression**. The American Statistician, v. 46, n. 3, p. 175-185, 1992.

ANDERSON, T.W.; DARLING, D.A. **Asymptotic Theory of Certain "Goodness-of-fit" Criteria Based on Stochastic Processes**. Em: Mathematical Statistics. p. 193–212, 1952

BECKER, H.; NAAMAN, M.; GRAVANO, L. **Learning similarity metrics for event identification in social media**. Em: The Third ACM international conference on Web search and data mining (WSDM '10). ACM, New York, NY, USA, p. 291-300, 2010

BECKER, H.; NAAMAN, M.; GRAVANO, L. **Selecting Quality Twitter Content for Events**. Em: The Fifth International AAAI Conference on Weblogs and Social Media (ICWSM '11). AAAI, Barcelona, Espanha, p. 442-445, 2011

BECKER, H; ITER, D; NAAMAN, M; GRAVANO, L., **Identifying Content for Planned Events Across Social Media Sites**. Em: The Fifth ACM International Conference on Web Search and Data Mining (WSDM '12). ACM, New York, NY, USA, p. 533-542, 2012

BOSLAUGH, S.; WATTERS, P. **Statistics in a Nutshell**. O'REILLY, 2008

BRENNER, M.; IZQUIERDO, E. **Social Event Detection and Retrieval in Collaborative Photo Collections**. Em: The 2nd ACM International Conference on Multimedia Retrieval (ICMR '12). ACM, Hong Kong, China, artigo 21, 8 páginas, 2012

BRENNER, M.; IZQUIERDO, E. **MEDIAEVAL 2013: Social Event Detection, Retrieval and Classification in Collaborative Photo Collections**. Em: MEDIAEVAL '13. 2013

CHOUDHURY, M. D.; DIAKOPOULOS, N.; NAAMAN, M. **Unfolding the Event Landscape on Twitter: Classification and Exploration of User Categories**. Em: The ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12). ACM, New York, NY, USA, p. 241-244, 2012

COOPER, M.; FOOTE, J.; GIRGENSOHN, A.; WILCOX, L. **Temporal Event Clustering for Digital Photo Collections**. Em: ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), v. 1, n. 3, p. 269–288, 2005.

CP-3451, J. **Exchangeable image file format for digital still cameras: Exif Version 2.2**. Japan Electronics and Information Technology Industries Association, 2002.

DAO, M.; BOATO, G.; NATALE, F. G.B. DE; NGUYEN, T. **Jointly Exploiting Visual and Non-visual Information for Event-related Social Media Retrieval**. Em: The 3rd ACM International Conference on Multimedia Retrieval (ICMR '13). ACM, New York, NY, USA, p. 159-166.

DEY, A.K.; ABOWD, G.D. Towards a Better Understanding of Context and Context Awareness. Em: CHI 2000 Workshop of Context-Awareness. 2000.

ESTER, M.; KRIEGEL, H.; SANDER, J.; XU, X.; SIMOUDIS, E.; HAN, J.; FAYYAD, U. M. **A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise**. Em: The Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI, p. 226–231, Portland, Oregon, 1996

FERREIRA, A. D. H. **Novo dicionário Aurélio da língua portuguesa**. Editora Positivo. 2004.

FIGUEIRÊDO, H. F. DE; LACERDA, Y. A.; PAIVA, A. C. DE; CASANOVA, M. A.; BAPTISTA, C. S. **PhotoGeo: A Photo Digital Library with Spatial-temporal Support and Self-annotation**. Em: Multimedia Tools and Applications. Springer, v. 59, n. 1, p. 279–305, 2012.

GAO, H.; TANG, J.; LIU, H. **Mobile Location Prediction in Spatio-temporal Context**. Em: Nokia Mobile Data Challenge 2012 Workshop. v. 2, n. 1, 2012.

GAO, X.; CAO, J.; JIN, Z.; LI, X.; LI, J. **GESODECK: A Geo-social Event Detection and Tracking System**. Em: The 21st ACM International Conference on Multimedia (MM '13). ACM, New York, NY, USA, p.471-472, 2013

GONG, Y.; LI, Y.; JIN, D.; SU, L.; ZENG, L. **A Location Prediction Scheme Based on Social Correlation**. Em: Vehicular Technology Conference (VTC SPRING), IEEE, p. 1-5, 2011.

GRABOVITCH-ZUYEV, I.; KANZA, Y.; KRAVI, E.; PAT, B. **On the Correlation Between Textual Content and Geospatial Locations in Microblogs**. Em: The First International ACM Workshop on Managing and Mining Enriched Geo-spatial Data (GEORICH '14). ACM, 2014

HARTIGAN, J. A. **Clustering Algorithms**. 1975.

HAYS, J.; EFROS, A. A. **IM2GPS: Estimating Geographic Information from a Single Image**. Em: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008

HOLLENSTEIN, L.; PURVES, R. **Exploring Place Through User-generated Content: Using Flickr Tags to Describe City Cores**. Em: Journal of Spatial Information Science. p. 21-48, 2010.

ILINA, E.; HAUFF, C.; CELIK, I.; ABEL, F.; HOUBEN, G. **Social Event Detection on Twitter**. Em: 12th International Conference on Web Engineering (ICWE '12). Springer, p. 169-176, 2012

IVANOV, I.; VAJDA, P.; LEE, J. S.; GOLDMANN, L.; EBRAHIMI, T. **Geotag Propagation in Social Networks Based on User Trust Model**. Em: Multimedia Tools and Applications. Springer, p. 155-177, 2012.

PANIAGUA, J., TANKOYEU, I., STÖTTINGER, J.; GIUNCHIGLIA, F. **Social Events and Social Ties**. Em: The 3rd ACM International Conference on Multimedia Retrieval (ICMR '13). ACM, New York, NY, USA, p. 143-150, 2013.

JOSHI, D.; GALLAGHER, A.; YU, J.; LUO, J. **Inferring Photographic Location Using Geotagged Web Images**. Em: Multimedia Tools and Applications. Springer, v. 56, n. 1, p. 131–153, 2012.

FENG, K.; CONG, G.; BHOWMICK, S. S.; MA, S. **In Search of Influential Event Organizers in Online Social Networks**. Em: The 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD '14). ACM, New York, NY, USA, p.63-74, 2014.

FREEDMAN, D. **Statistical models: theory and practice**. Cambridge University Press, 2009.

LACERDA, Y. A.; DE FIGUEIRÊDO, H. F.; DE SOUZA BAPTISTA, C.; SAMPAIO, M. C. **PhotoGeo: A Self-organizing System for Personal Photo Collections**. Em: The 10th IEEE International Symposium on Multimedia. IEEE, p. 258–265, 2008.

LACERDA, Y. A.; FIGUEIRÊDO, H. F. DE; SILVA, J. P. R. DA; LEITE, D. F. B.; PAIVA, A. C. DE; BAPTISTA, C. DE S. **On Improving Geotag Quality in Photo Collections**. Em: The Fifth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOPROCESSING 2013). IARIA, NICE, FRANCE, 2013.

LIM, J. H.; TIAN, Q.; MULHEM, P. **Home Photo Content Modeling for Personalized Event-based Retrieval**. Em: IEEE Multimedia. IEEE, v. 10, n. 4, p. 28–37, 2003.

MATELLANES, A.; EVANS, A.; EEDAL, B., **Creating an application for automatic annotations of images and video**. Em: 1st International Workshop on Semantic Web Annotations for Multimedia (SWAMM). Edinburgh, Scotland, 2006.

MEZARIS, V. ; SCHERP, A.; JAIN, R.; KANKANHALLI, M., **Real-life Events in Multimedia: Detection, Representation, Retrieval, and Applications**. Em: Multimedia Tools and Applications. Springer, v. 70, n. 1, p. 1-6, 2014.

MEYER, C. D. **Matrix analysis and applied linear algebra**. Siam, 2000.

MITCHELL, T. M. **Machine Learning**. McGraw-Hill Inc., New York, NY, USA, 1 edition, 1997.

NAAMAN, M.; HARADA, S.; WANG, Q.; GARCIA-MOLINA, H.; PAEPCKE, A. **Context Data in Geo-referenced Digital Photo Collections**. Em: 12th Annual ACM International Conference on Multimedia. ACM, p. 196–203, 2004.

NAAMAN, M.; NAIR, R. **ZoneTag's Collaborative Tag Suggestions: What is This Person Doing in My Phone?** Em: IEEE MULTIMEDIA. IEEE, v. 15, n. 3, p. 34–40, 2008.

NAKAJI, Y.; YANAI, K. **Visualization of Real-world Events with Geotagged Tweet Photos**. Em: IEEE International Conference on Multimedia and Expo Workshops. IEEE, p.272–277, 2012.

The MEDIAEVAL 2013 Multimedia Benchmark Workshop, Barcelona, Espanha, 2013. **Proceedings...** CEUR-WS.ORG. 2013

PSALLIDAS, F.; BECKER, H.; NAAMAN, M.; GRAVANO, L. **Effective Event Identification in Social Media**. Em: IEEE Data Eng. IEEE, p. 42-50, 2013.

RABBATH, M.; SANDHAUS, P.; BOLL, S. **Analysing Facebook Features to Support Event Detection for Photo-based Facebook Applications**. Em: The 2nd ACM International Conference on Multimedia Retrieval (ICMR '12), ACM, Hong Kong, China, p. 1–8, 2012.

RAJARAMAN, A.; ULLMAN, J. D. **Mining of Massive Datasets**. Cambridge University, 2011.

RIJSBERGEN, C. J. **Information Retrieval**. 1979.

REUTER, T.; PAPADOPOULOS, S.; MEZARIS, V.; CIMIANO, P. **RESEED: Social Event Detection Dataset**. Em: The 5th ACM Multimedia Systems Conference (MMSYS '14). ACM, New York, NY, USA, p. 35-40, 2014.

RODDEN, K.; WOOD, K. R. **How do People Manage Their Digital Photographs?** Em: SIGCHI Conference on Human Factors in Computing Systems. ACM, p. 409–416.

SARVAS, R.; HERRARTE, E.; WILHELM, A.; DAVIS, M., **Metadata creation system for mobile images**. Em: 2th International Conference on Mobile Systems, Applications, and Services (MobiSys '04). ACM, Boston, MA, USA, p 36-48, 2004.

SHAPIRO, S. S.; WILK, M. B. **An Analysis of Variance Test for Normality**. Biometrika, v. 52, n. 3/4, p.591–611, 1965.

SHIMIZU, K.; NITTA, N.; BABAGUCHI, N. **Learning People Co-occurrence Relations by Using Relevance Feedback for Retrieving Group Photos**. Em: The 1st ACM International Conference on Multimedia Retrieval. ACM, artigo 57, 8 páginas, 2011.

STONE, Z.; ZICKLER, T.; DARRELL, T. **Toward Large-scale Face Recognition Using Social Network Context**. Em: The IEEE. IEEE, v. 98, n. 8, p. 1408–1415, 2010.

TREFETHEN, L. N.; BAU III, D. **Numerical linear algebra**. Siam, 1997.

VIANA, W.; BRINGEL FILHO, J.; GENSEL, J.; OLIVER, M. V.; MARTIN, H. **PhotoMap–Automatic Spatiotemporal Annotation for Mobile Photos**. Em: Web and Wireless Geographical Information Systems. Springer, p. 187-201, 2007.

VYAS, D.; NIJHOLT, A.; VEER, G. V. D. **Practices Surrounding Event Photos**. Em: Human-computer Interaction - INTERACT. Springer, p.55-72, 2013.

ZIGKOLIS, C.; PAPADOPOULOS, S.; FILIPPOU, G.; KOMPATSIARIS, Y.;
VAKALI, A. **Collaborative Event Annotation in Tagged Photo Collections**. Em:
Multimedia Tools and Applications. Springer, p. 1–30, 2012.