



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ANDERSON ALMEIDA FIRMINO

**Anotação automática e semiautomática de pessoas em
fotografias digitais utilizando eventos compartilhados**

DISSERTAÇÃO

Campina Grande
Agosto de 2017

ANDERSON ALMEIDA FIRMINO

Anotação automática e semiautomática de pessoas em fotografias digitais utilizando eventos compartilhados

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.

Orientador: Cláudio de Souza Baptista

Campina Grande
Agosto de 2017

DIGITALIZAÇÃO:
SISTEMOTECA - UFCG

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

F525a Firmino, Anderson Almeida.
 Anotação automática e semiautomática de pessoas em fotografias digitais utilizando eventos compartilhados / Anderson Almeida Firmino. – Campina Grande, 2017.
 84 f. : il. color.

 Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2017.
 "Orientação: Prof. Dr. Cláudio de Souza Baptista".
 Referências.

 1. Anotação de Pessoas em Fotografias. 2. Detecção de Eventos. 3. Coleção de Fotografias Pessoais – Anotação Automática. I. Baptista, Cláudio de Souza. II. Título.

CDU 004.021(043)

**"ANOTAÇÃO AUTOMÁTICA E SEMIAUTOMÁTICA DE PESSOAS EM FOTOGRAFIAS
DIGITAIS UTILIZANDO EVENTOS COMPARTILHADOS"**

ANDERSON ALMEIDA FIRMINO

DISSERTAÇÃO APROVADA EM 02/08/2017


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


EANES TORRES PEREIRA, Dr., UFCG
Examinador(a)


HUGO FEITOSA DE FIGUEIRÊDO, Dr., IFPB
Examinador(a)

WINDSON VIANA DE CARVALHO, Dr., UFC
Examinador(a)

CAMPINA GRANDE - PB

Agradecimentos

A Deus em primeiro lugar, pois se não fosse por Ele, eu não teria chegado até aqui. Ao Pai, pelo encorajamento e fortalecimento nos momentos difíceis. Ao Filho, pelo socorro bem presente na angústia. Ao Espírito Santo, pelo ensino em tempos de dúvidas. Obrigado Deus por tudo!

À minha família pelo apoio em todo o tempo. Aos meus irmãos, Manassés e Misael; ao meu pai - Manoel - e, em especial, à minha mãe Lúcia, por sempre ter acreditado em mim e me incentivado.

Ao grupo Dunamis, pelas orações, pelos momentos de alegria compartilhados e pelo incentivo. Em especial, à missionária Carmem, por todo o suporte e ajuda em momentos difíceis.

Ao meu orientador, Prof. Dr. Cláudio de Souza Baptista, pela orientação e paciência, por ter me convidado para ingressar no mestrado e por todo o aprendizado transmitido. Obrigado, Professor!

Ao meu coorientador, Prof. Dr. Hugo Feitosa de Figueirêdo, por toda a ajuda e acompanhamento ao longo da realização desta pesquisa.

Aos colegas do LSI, pelos momentos de aprendizado e por proporcionarem um ótimo ambiente de trabalho. Em especial, a Gabriel Joseph, pela ajuda e contribuição direta na realização desta pesquisa.

Aos colegas do LVC, nas pessoas do Prof. Dr. Eanes Torres e Ronnyldo Silva, pela contribuição na realização desta pesquisa.

À UFCG e ao LSI, por minha formação acadêmica.

Resumo

Diante da popularização de smartphones e dispositivos fotográficos, é fato que as fotografias digitais têm se tornado parte do cotidiano. Milhões de fotos são adicionadas diariamente em redes sociais, como Facebook e Instagram. Diante de isso, tem-se a seguinte questão: "Como gerenciar tamanho acervo de fotos?" Para o usuário, isto não é uma tarefa simples. Para auxiliá-lo na resolução dessa questão, se faz necessário o uso de anotações em fotografias, as quais podem conter informações de localização, data, hora e *tags* de conteúdo. Para o usuário, a identificação das pessoas presentes na cena é considerada a informação mais importante, permitindo que ele se recorde da fotografia. Por outro lado, utilizar o conceito do evento a qual uma fotografia pertence é uma forma de otimizar a organização de fotografias. Neste trabalho, propõe-se verificar se o conceito de evento compartilhado, o qual consiste em diversas pessoas com smartphones e dispositivos fotográficos participando de um mesmo evento, traz algum incremento na taxa de acerto na anotação de pessoas em fotografias digitais de maneira automática e semiautomática. Para a anotação automática de pessoas, foram utilizados algoritmos de detecção e reconhecimento de faces; e para a anotação semiautomática de pessoas, foram utilizados estimadores baseados nas informações de contexto das fotografias (como localização geográfica, por exemplo) e de conteúdo (como as faces detectadas, por exemplo). Os experimentos realizados apresentaram bons resultados, demonstrando assim que a utilização do conceito de evento compartilhado aumenta a taxa de acerto para a anotação automática e para a anotação semiautomática de pessoas em fotografias digitais.

Abstract

Faced with the popularization of smartphones and photographic devices, it is a fact that digital photographs have become part of daily life. Millions of photos are added daily on social networks such as Facebook and Instagram. Faced with this, one has the following question: "How to manage huge photo collections?" For the user, this is not a simple task. To assist in resolving this issue, it is necessary to use photo annotations, which may contain location information, date, time, and content tags. For the user, identifying the people present in the scene is considered the most important information, allowing him to remember the photograph. On the other hand, using the concept of the event to which a photograph belongs is a way of optimizing the organization of photographs. In this work, it is proposed to verify if the concept of shared event, which consists of several people with smartphones and photographic devices participating in the same event, brings some increase in the hit rate in annotating people in digital photographs in an automatic and semi-automatic way. For the automatic people annotation, we used algorithms of face detection and recognition; and for the semi-automatic people annotation, estimators based on the context information of the photographs (such as geographic location, for example) and content (such as detected faces, for example) were used. The experiments performed showed good results, thus demonstrating that the use of the concept of shared event increases the hit rate for both people automatic and semiautomatic annotation of in digital photographs.

Lista de ilustrações

Figura 1 – Exemplo de evento compartilhado	12
Figura 2 – Exemplo de metadados no formato EXIF	16
Figura 3 – Exemplo de agrupamento utilizando o algoritmo DBSCAN	20
Figura 4 – Exemplos de faces detectadas utilizando o algoritmo de Viola e Jones (2001)	27
Figura 5 – Exemplos de faces detectadas utilizando o algoritmo proposto por Markus et al. (2013)	28
Figura 6 – Comparação de detectores de face utilizados na anotação automática de pessoas	29
Figura 7 – Comparação de detectores de face utilizados na anotação semiautomática de pessoas	30
Figura 8 – Visão geral do modelo	44
Figura 9 – Exemplo de cálculo de similaridade entre dois eventos	49
Figura 10 – À esquerda (quadro vermelho) estão as faces detectadas pelo algoritmo; à direita (quadro verde) estão as faces mais semelhantes às respectivas faces detectadas.	54
Figura 11 – Solução proposta por Figueirêdo et al., 2015	55
Figura 12 – Exemplos de imagens da base de dados de Gallagher e Chen (2008)	58
Figura 13 – Exemplos de imagens da base de dados de Gallagher e Chen (2009)	59
Figura 14 – Exemplos de imagens da base de dados de Conci et al. (2015) . . .	60
Figura 15 – Comparação da execução do algoritmo de detecção e reconhecimento de faces usando uma base pública e a base de fotos pessoais adotada	66
Figura 16 – Comparação de resultados de anotação automática de pessoas sobre o uso de eventos compartilhados	67
Figura 17 – Resultados da anotação semiautomática de pessoas considerando eventos compartilhados	68
Figura 18 – Resultados detalhados da anotação semiautomática de pessoas considerando eventos compartilhados	69

Lista de tabelas

Tabela 1 – Comparação dos trabalhos relacionados	38
Tabela 2 – Resultados detalhados da comparação entre anotação semiautomática de pessoas utilizando ou não o conceito de evento compartilhado	70
Tabela 3 – Resultados do teste de normalidade para os dados da anotação automática de pessoas	82
Tabela 4 – Resultados do teste de hipóteses concernente a eventos compartilhados e anotação automática de pessoas	83
Tabela 5 – Resultados do teste de normalidade para os dados da anotação semiautomática de pessoas	83
Tabela 6 – Resultados do teste de hipóteses concernente a eventos compartilhados e anotação semiautomática de pessoas	84

Sumário

1	INTRODUÇÃO	10
1.1	Objetivos	12
1.2	Hipóteses	13
1.3	Contribuições	13
1.4	Publicações	13
1.5	Organização Estrutural	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Anotações em fotografias digitais	15
2.1.1	Pessoas	16
2.1.2	Eventos	17
2.2	Algoritmos	19
2.2.1	Detecção de eventos pessoais	19
2.2.1.1	DBSCAN	19
2.2.2	Detecção de eventos compartilhados	21
2.2.3	Anotação semiautomática de pessoas em fotografias	22
2.2.3.1	Estimadores baseados em contexto	22
2.2.3.2	Estimadores baseados em conteúdo	23
2.2.3.3	Geração da lista de recomendações	24
2.2.3.4	Ponderação dos estimadores	24
2.2.3.4.1	Regressão linear múltipla	25
2.2.4	Detecção de Faces	26
2.3	Considerações Finais	30
3	TRABALHOS RELACIONADOS	31
3.1	Anotações de pessoas	31
3.2	Anotações de eventos	34
3.3	Comparação dos trabalhos relacionados	38
3.4	Considerações Finais	42
4	ANOTAÇÃO DE PESSOAS UTILIZANDO EVENTO COMPARTILHADO	43
4.1	Detecção de Eventos Compartilhados	45
4.1.1	Detecção de Eventos Pessoais	46
4.1.2	Detecção de Eventos Compartilhados	47
4.2	Anotação Automática de Pessoas em Fotografias	49

4.2.1	Detecção de faces	50
4.2.2	Reconhecimento de faces	50
4.3	Anotação Semiautomática de Pessoas em Fotografias	55
4.4	Considerações Finais	56
5	AVALIAÇÃO E RESULTADOS	57
5.1	Bases de dados utilizadas	57
5.2	Metodologia de Avaliação	62
5.2.1	Anotação automática de pessoas em fotografias	63
5.2.2	Anotação semiautomática de pessoas em fotografias	64
5.3	Resultados	65
5.3.1	Anotação automática de pessoas em fotografias	65
5.3.2	Anotação semiautomática de pessoas em fotografias	68
5.4	Considerações Finais	71
6	CONCLUSÕES E TRABALHOS FUTUROS	72
6.1	Contribuições	73
6.2	Trabalhos Futuros	73
	REFERÊNCIAS	75
	APÊNDICES	81
	APÊNDICE A – TESTES ESTATÍSTICOS	82
A.1	Anotação automática de pessoas	82
A.2	Anotação semiautomática de pessoas	83

1 Introdução

Uma pesquisa recente realizada pelo grupo Gartner mostra que a venda de smartphones ultrapassou um bilhão e quatrocentos milhões de unidades em 2015¹. Diante da popularização destes aparelhos, bem como das câmeras digitais, é inegável o fato de que as fotografias digitais têm se tornado parte do cotidiano. De acordo com dados atualizados do Facebook, mais de 300 milhões de fotos são adicionadas diariamente². Já com relação ao Instagram, tem-se que cerca de 80 milhões de fotos e vídeos são adicionados diariamente³.

Gerenciar e organizar grandes quantidades de fotos são tarefas extenuantes para o usuário. Para auxiliar na resolução desse problema, a utilização de anotações em fotografias torna-se imprescindível. Assim, as fotos podem ser vinculadas a metadados, que contêm informações como localização, hora, data e tags de conteúdo (HANBURY, 2008). A maioria das ferramentas de gerenciamento de fotografias (por exemplo, iPhoto, Picasa, Google Fotos) utiliza metadados como entrada para métodos de recomendação ou inferência de anotações de fotografias.

Estudos mostram que a anotação de informações sobre o contexto na qual a fotografia foi capturada auxilia em sua recuperação futura. De acordo com Ames e Naaman (2007), anotações podem melhorar significativamente a usabilidade de coleções de fotos, facilitando a realização de pesquisas em tais coleções. Mais especificamente, a identificação das pessoas presentes na cena é considerada a informação mais importante, permitindo que uma pessoa se recorde da fotografia (NAAMAN et al., 2005). No entanto, a maior parte das fotografias digitais não são anotadas com tal informação.

O processo de anotação de pessoas em fotografias pode ser feito de três diferentes maneiras: manualmente, na qual o usuário realiza a anotação sem o auxílio de uma aplicação; semiautomaticamente, na qual uma aplicação ajuda o usuário a realizar a anotação; e automaticamente, onde uma aplicação anota as fotografias, podendo se utilizar de outras fontes e de técnicas baseadas em conteúdo (ESCALANTE; MONTES-Y-GÓMEZ; SUCAR, 2012).

Nos últimos anos, técnicas de organização de fotos automática utilizando anotações têm sido propostas. Existem várias abordagens que utilizam aspectos, como as pessoas presentes nas fotos (SHIMIZU; NITTA; BABAGUCHI, 2011; STONE; ZICKLER; DARRELL, 2008), para ajudar na organização e recuperação destas.

¹ <http://www.gartner.com/newsroom/id/3215217>

² <https://zephoria.com/top-15-valuable-facebook-statistics/>

³ <https://searchenginewatch.com/2016/04/20/23-stats-and-facts-about-instagram/>

Muitas abordagens na literatura abordam o uso de técnicas semiautomáticas para anotar pessoas em fotografias (NAAMAN et al., 2005; O'HARE; SMEATON, 2009). As técnicas que são baseadas nas informações contextuais das fotos geralmente usam estimadores para gerar uma recomendação das pessoas que estão mais propensas a estarem em uma certa fotografia. Tais estimadores são estatísticas baseadas em informação contextual de fotografias previamente capturadas, a exemplo de recorrência espacial de um contato em uma determinada área e recorrência geral de um contato em uma coleção de fotografias. Em geral, a combinação desses estimadores gera bons resultados.

Alguns pesquisadores da área perceberam que uma das maneiras adotadas para otimizar a organização de fotos é utilizando o conceito de evento (RODDEN; WOOD, 2003). Eles seguem abordagens distintas em se tratando de organizar fotos utilizando eventos. Alguns estudos sugerem anotar os eventos nas fotografias (RABBATH; SANDHAUS; BOLL, 2012; DAO et al., 2013), enquanto outros sugerem a clusterização de fotos do mesmo evento (RUOCCO; RAMAMPIARO, 2015), independentemente da anotação de eventos. Estas abordagens podem incluir dois tipos de eventos: pessoais e sociais (BRENNER; IZQUIERDO, 2012; ZIGKOLIS et al., 2014). Eventos pessoais estão relacionados à vida pessoal de um indivíduo - aniversário da vovó, por exemplo. Já eventos sociais representam um evento na sociedade, a exemplo do Festival Pop Monterey.

Por outro lado, um conceito comumente desconsiderado no momento de detectar os eventos é o de evento compartilhado (JANG; YOON; CHO, 2009a; FIGUEIRÉDO et al., 2012; ANDRADE et al., 2014), que ocorre devido à existência de vários dispositivos (câmeras ou smartphones) de pessoas distintas em um mesmo evento. Como consequência disto, é possível que fotografias capturadas num mesmo evento possuam tipos de informações diferentes nos metadados, que podem ser compartilhados entre as pessoas presentes, gerando uma espécie de anotação cruzada.

A figura 1 ilustra um exemplo de evento compartilhado. Nela, pode-se ver que os usuários Ana e José compareceram a um evento em comum - o casamento de Pedro e Maria. Após o término do evento, Ana e José enviam as fotos capturadas no evento para suas respectivas redes sociais. No entanto, sem levar em consideração o conceito de evento compartilhado, as pessoas anotadas por Ana não são utilizadas na coleção de José e vice-versa. Com a utilização deste conceito, as anotações podem ser compartilhadas entre Ana e José.

Figura 1 – Exemplo de evento compartilhado



Fonte: Elaborado pelo autor

Andrade et al. (2014) abordaram o conceito de evento compartilhado em conjunto com a anotação da localização geográfica em fotografias. Os autores detectaram eventos compartilhados e realizaram a propagação da anotação da localização geográfica para cada evento.

Figueirêdo et al. (2012) já trataram do conceito de evento compartilhado anteriormente. No entanto, a abordagem deles demandava que as fotografias estivessem previamente anotadas com a localização geográfica e com as pessoas presentes nas fotos. Sabe-se que a maior parte das fotografias não são anotadas com a informação das pessoas presentes.

Neste sentido, o intuito deste trabalho é melhorar a anotação automática e semiautomática de pessoas em fotografias digitais, utilizando o conceito de evento compartilhado.

1.1 Objetivos

Considerando-se o contexto de organização de coleções de fotografias e a importância de pessoas e dos eventos relacionados às fotografias, este trabalho tem como objetivo propor uma metodologia para realizar anotações automáticas e semiautomáticas de pessoas utilizando o conceito de eventos compartilhados.

Os objetivos específicos deste trabalho são:

- O1: Elaborar algoritmo de anotação semiautomática de pessoas utilizando o conceito de evento compartilhado;
- O2: Elaborar algoritmo de anotação automática de pessoas utilizando o conceito de evento compartilhado;

1.2 Hipóteses

As seguintes hipóteses foram consideradas e testadas no escopo da pesquisa descrita nesta dissertação:

- H1: A utilização do conceito de evento compartilhado melhora a taxa de acerto para anotação automática de pessoas;
- H2: A utilização do conceito de evento compartilhado melhora a taxa de acerto para anotação semiautomática de pessoas.

1.3 Contribuições

Neste trabalho, as principais contribuições são:

- A proposição de uma técnica para realizar anotações automáticas de pessoas utilizando o conceito de evento compartilhado;
- A proposição de uma técnica para realizar anotações semiautomáticas de pessoas utilizando o conceito de evento compartilhado;

1.4 Publicações

Esta pesquisa resultou no seguinte trabalho submetido:

- Automatic and semiautomatic annotation of people on photographs using shared event - submetido ao Expert Systems With Applications, 2017.

1.5 Organização Estrutural

O restante desta dissertação está organizado da seguinte forma: no capítulo 2, é apresentada a fundamentação teórica, que aborda os fundamentos da anotação de pessoas e eventos em fotografias. No capítulo 3, são mostrados os trabalhos relacionados a esta pesquisa. No capítulo 4, é descrita a proposta para anotar pessoas em fotografias de forma automática e semiautomática utilizando o conceito de evento compartilhado. No capítulo 5, a metodologia de avaliação utilizada e os resultados

obtidos são detalhados. Finalmente, no capítulo 6, estão descritas as conclusões e os trabalhos futuros.

2 Fundamentação Teórica

O objetivo deste capítulo é abordar os principais conceitos utilizados nesta pesquisa. Para tanto, ele está dividido em três seções, que discorrem sobre anotações em fotografias digitais e algoritmos de classificação e agrupamento. Ao fim, algumas considerações são apresentadas.

2.1 Anotações em fotografias digitais

Devido ao grande número de fotos que estão sendo geradas atualmente, é muito importante ter técnicas que permitam a organização e recuperação dessas imagens. A anotação em fotografias digitais desempenha um papel fundamental neste sentido, pois pode vincular dados brutos (isto é, as fotografias) a informações específicas que auxiliam os seres humanos a lidar com grandes quantidades de conteúdo.

De acordo com Andrade et al. (2016), a anotação de fotografias pode ser entendida como o processo pelo qual um sistema computacional ou uma pessoa atribui metadados sob a forma de legendas ou palavras-chave à uma fotografia digital. Estes metadados contêm informações como localização, hora, data e tags de conteúdo.

O processo de anotação em fotografias digitais pode ser feito com base no conteúdo da imagem e/ou no contexto em que a imagem está inserida. No escopo deste trabalho, considera-se que o conteúdo da imagem são as informações que podem ser extraídas a partir de um processamento visual efetuado nas fotografias, enquanto o contexto é definido como algo que descreve a fotografia em si, a exemplo de localização, data e hora da captura da fotografia.

Uma fonte importante de informação das imagens é o formato EXIF (EXchangeable Image File). É uma forma de armazenar metadados que está presente na maioria das câmeras fotográficas do mercado. Esse formato contém informações como nome, tamanho, data/hora, modelo da câmera, coordenadas geográficas. Um exemplo de informações armazenadas no tipo de formato EXIF pode ser visto na figura 2

No entanto, fazer com que palavras chave sejam atribuídas a imagens por meio de sistemas computacionais não é uma tarefa simples. Ivasic-Kos, Ipsic e Ribaric (2015) desenvolveram uma abordagem utilizando representação de conhecimento baseado em Redes de Petri para representar o conhecimento sobre conceitos que podem aparecer em uma imagem e, assim, realizar anotações automáticas de diferentes níveis de abstração nas fotografias, identificando objetos cenários e conceitos associados a elas.

Figura 2 – Exemplo de metadados no formato EXIF

EXIF Information			
File name:	DSC_0260.JPG	File size:	922866 bytes
File date:	2006:04:22 22:06:16	Camera make:	NIKON CORPORATION
Camera model:	NIKON D70s	Date/Time:	2006:04:17 18:06:08
Resolution:	3000 x 2632	Flash used:	No
Focal length:	18.0mm (35mm equivalent: 27mm)	Exposure time:	0.0008 s (1/1250)
Aperture:	f/8.0	Whitebalance:	Manual
Metering Mode:	matrix	Exposure:	Manual
Exposure Mode:	ManualAuto bracketing		

Fonte: <http://direitoeti.com.br/artigos/exif-metadata-a-investigacao-policial-subsidiada-por-sua-extracao-e-analise/> (2017)

O uso de anotações permite que ferramentas de gerenciamento de fotografias sejam aprimoradas e diminuam o trabalho do usuário final. Como exemplo, pode-se citar o fato de que a anotação da localização geográfica a qual a fotografia pertence permite uma organização direta em relação ao local e uma apresentação da fotografia em um mapa. Assim, é possível visualizar fotografias capturadas próximas a um determinado local, desde que sejam fornecidos o nome ou as coordenadas do lugar.

2.1.1 Pessoas

O processamento de imagens em conjunto com técnicas de visão computacional resulta em informações sobre a imagem, a exemplo de se a captura ocorreu em ambiente aberto ou fechado. Informações deste tipo são úteis para descrever o contexto no qual a fotografia foi capturada. A evolução das câmeras e smartphones vem inserindo esse processamento digital diretamente nos dispositivos.

As técnicas para realizar anotações de pessoas presentes nas fotos podem ser divididas em três grupos: técnicas que utilizam reconhecimento de faces; técnicas que são baseadas na utilização de estimadores calculados com base em padrões das informações contextuais relacionadas ao ambiente no qual a fotografia foi capturada; e técnicas híbridas, as quais melhoram o desempenho do reconhecimento de faces usando os estimadores das informações de contexto (FIGUEIRÉDO et al., 2015).

Para a realização de anotação manual de pessoas em fotografias é necessário um esforço extenuante por parte do usuário, ocasionando a não realização de ano-

tações pela maioria dos donos de grandes coleções de fotografias, muito embora as pessoas presentes nas fotografias sejam as principais informação utilizada em buscas.

Uma solução possível para este problema seria a anotação automática das pessoas presentes nas fotografias. No entanto, esta solução tem como limitação o baixo desempenho de técnicas de anotações automáticas, as quais são baseadas no reconhecimento de faces, em geral.

Outra solução adotada na literatura é a anotação semiautomática de pessoas nas fotografias, a qual consiste na geração de uma lista de sugestões com as pessoas mais prováveis de estarem presentes nas fotografias. Utilizando esta solução, o usuário apenas precisa confirmar as pessoas que estão de fato presentes nas fotos, sem a necessidade de anotá-las manualmente.

Algumas soluções para anotar pessoas em fotografias de forma semiautomática (NAAMAN et al., 2005; FIGUEIRÊDO et al., 2015) utilizaram o conceito de estimadores, que são estatísticas baseadas em uma ou mais informações contextuais das fotografias capturadas. Alguns exemplos de estimadores são: recorrência geral de um contato na coleção de fotografias e recorrência geográfica de um contato em uma determinada área. Em geral, a combinação desses estimadores gera bons resultados.

Em se tratando de técnicas automáticas de anotações de pessoas que utilizam reconhecimento facial, é inegável que tais técnicas auxiliam na realização de anotações em fotografias. No entanto, estas técnicas são limitadas por apresentarem baixa taxa de acerto (PHAM et al., 2007).

Utilizar apenas técnicas de reconhecimento de faces para anotar pessoas pode produzir resultados insatisfatórios. Isso se dá devido ao fato de as faces nem sempre estarem iluminadas ou serem capturadas adequadamente nas fotografias. Além disso, grandes variações de pose ou expressão e oclusões parciais (como o uso de óculos de sol ou maquiagem) são comuns e difíceis de lidar (BRENNER; MIRZA; IZQUIERDO, 2014).

Deste modo, utilizar o reconhecimento de faces em conjunto com uma abordagem de anotação semiautomática, que utiliza informações contextuais, torna-se uma solução mais completa para o usuário, dado que este poderá confirmar a anotação das pessoas que não foram detectadas pelo reconhecimento de faces a partir da lista de sugestões.

2.1.2 Eventos

Naaman et al. (2004) pontuaram que além das pessoas presentes nas fotografias, outros atributos que ajudam na organização das fotografias são “Onde?” e “Quando?”. Com isto, pode-se entender que o evento no qual uma fotografia foi cap-

turada é um conceito de grande relevância no tocante a organização de coleções de fotografias, podendo ser utilizado para aprimorar a recuperação delas. Deste modo, automatizar a organização de fotografias em eventos resulta na diminuição dos esforços empregados em uma incumbência que vem sendo realizada pelas pessoas.

Um evento pode ser definido como um acontecimento que reúne pessoas com algum objetivo em comum durante um período de tempo e em um determinado lugar. Exemplos de eventos incluem festas de aniversário, shows e casamentos.

Segundo Viana et al. (2007), informações como data, hora e localização geográfica quando usadas puramente não são determinantes para que um usuário se recorde de uma determinada fotografia. Assim, depreende-se que o conceito de evento pode ser utilizado para preencher esta lacuna, permitindo que um usuário recupere fotografias de forma mais eficiente.

A detecção automática de eventos pode ser realizada através da utilização de informações contextuais das fotografias e fontes externas, como, por exemplo, as redes sociais. As principais informações contextuais das fotografias utilizadas para a detecção automática de eventos são: espacial (ex: localização geográfica) e temporal (ex: data de captura).

Alguns métodos existentes na literatura que realizam a detecção de eventos utilizam a ideia de que um evento pode possuir características hierárquicas, podendo assim existir sub-eventos de um evento maior (ZIGKOLIS et al., 2014). Por exemplo, “brincadeira na neve” e “preparação para o jantar” podem ser considerados sub-eventos de um evento maior chamado “Natal de 2016”. Desta forma, usuários diferentes podem possuir diferentes preferências no momento de selecionar em qual evento uma determinada fotografia foi capturada.

Por outro lado, um fator importante, denominado de evento compartilhado, tem sido comumente desconsiderado no momento de detectar os eventos. Esse fator ocorre devido à existência de vários dispositivos (câmeras ou smartphones) de pessoas distintas em um mesmo evento. Geralmente, cada pessoa possuidora de um dispositivo possui sua organização peculiar de fotografias. Como consequência do fato de dispositivos distintos possuírem funções distintas, é possível que fotografias capturadas num mesmo evento possuam tipos de informações nos metadados diferentes. Por exemplo, as fotografias capturadas a partir da câmera de um smartphone, equipada com GPS, possuem a informação de localização geográfica nos seus metadados, enquanto que em outra câmera, equipada com bússola, as fotografias capturadas possuem apenas as informações de orientação.

2.2 Algoritmos

Nesta seção, são explicados os algoritmos utilizados nesta pesquisa: detecção de eventos pessoais e compartilhados e os estimadores baseados em conteúdo e contexto.

2.2.1 Detecção de eventos pessoais

Um dos problemas mais pesquisados na literatura com relação à anotação de fotografias é a detecção automática de eventos. Nesta seção, é discorrido sobre o algoritmo de detecção de eventos usado, o qual foi proposto por Lacerda et al. (2008a).

Considerando E o conjunto de todos os eventos, $Y_{ev} \subseteq P$ o subconjunto que contém as fotografias capturadas no evento $ev \in E$, Cx_p as informações contextuais relacionadas com a fotografia $p \in P$ e Cn_p as informações de conteúdo relacionadas com a fotografia $p \in P$. A detecção de eventos pessoais consiste na geração de um subconjunto de fotografias $Y_{ev} \subseteq Y$, dado um conjunto de fotografias $Y \subseteq P$, de acordo com a Equação 2.1:

$$\omega: Co \rightarrow Pr \quad (2.1)$$

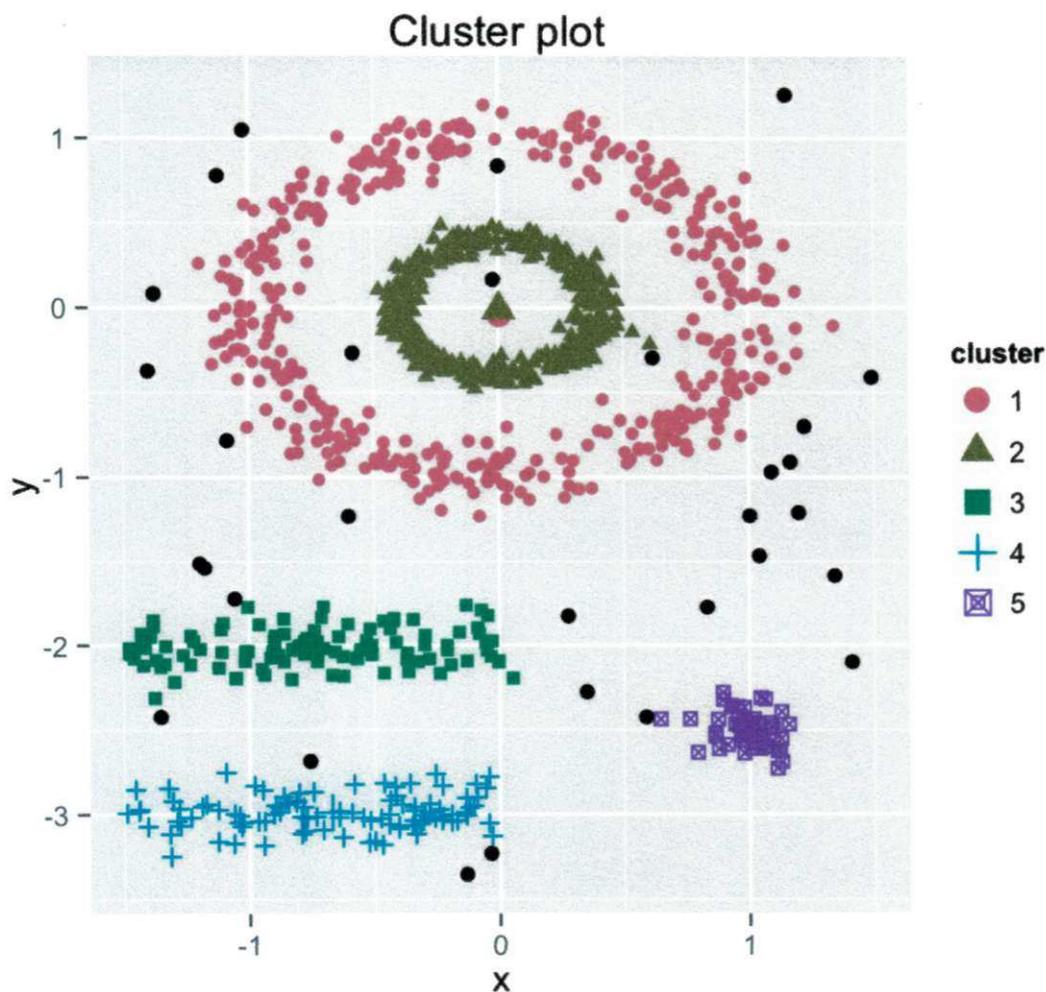
Esta função indica se duas fotografias p_i e p_{i+1} temporalmente consecutivas (ou espacialmente próximas) participam do mesmo evento, onde $Co = \{(p_i, p_{i+1}) \mid p_i \in P \text{ e } p_{i+1} \in P\}$ e $Pr = \{0, 1\}$. A fim de detectar todos os eventos, para cada fotografia $p_i \in Y$, a função ω é executada. Se $\omega(p_i, p_{i+1}) = 1$, então as fotos p_i e p_{i+1} pertencem ao evento Y_{ev} ($p_i, p_{i+1} \in Y_{ev}$). Caso contrário, se $\omega(p_i, p_{i+1}) = 0$, então $p_i \in Y_{ev}$, $p_{i+1} \notin Y_{ev}$ e $p_{i+1} \in Y_{ev+1}$ - ou seja, as fotografias p_i e p_{i+1} pertencem a eventos distintos.

2.2.1.1 DBSCAN

Para realizar o agrupamento de fotografias digitais em eventos, Lacerda et al. (2008a) utilizou o algoritmo DBSCAN (ESTER et al., 1996), que é um algoritmo de agrupamento espacial baseado em densidade de clusters. Assim, os clusters são definidos como regiões conectadas onde os pontos de dados são densos. Se a densidade de um ponto estiver abaixo de um determinado limiar, este é considerado como sendo ruído. Dado um conjunto de N pontos, DBSCAN produz uma partição a partir da entrada em um número de clusters e em um conjunto de pontos de ruído. O limite de densidade é especificado escolhendo o número mínimo de pontos (minPts) em uma esfera cujo raio é passado como parâmetro (epsilon). Como uma definição básica, um ponto central é um ponto cuja vizinhança satisfaz esta condição de densidade.

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: *epsilon* e *minPts*, onde o parâmetro *epsilon* indica a distância mínima entre dois pontos pertencentes a um mesmo grupo e o parâmetro *minPts* representa a quantidade mínima de pontos que cada grupo deve conter. Deste modo, o algoritmo encontrará os grupos adequados para a configuração de *epsilon* e *minPts* desejada.

Figura 3 – Exemplo de agrupamento utilizando o algoritmo DBSCAN



Fonte: <http://www.sthda.com/english/wiki/dbscan-density-based-clustering-for-discovering-clusters-in-large-datasets-with-noise-unsupervised-machine-learning> (2017)

Na figura 3, é apresentado um exemplo de agrupamento com o algoritmo DBSCAN. As áreas coloridas representam os clusters criados pelo algoritmo. Caso o algoritmo seja executado novamente, para a mesma base de dados e com as mesmas configurações de *epsilon* e *minPts*, os clusters serão sempre os mesmos. Porém, caso

os valores sejam modificados, certamente os grupos gerados serão diferentes.

2.2.2 Detecção de eventos compartilhados

Neste trabalho, foi utilizado o algoritmo de detecção de eventos compartilhados proposto por Figueirêdo et al. (2012). Após a criação de todos os clusters com eventos pessoais, a detecção de eventos compartilhados é realizada. Para isto, é verificada a similaridade entre todos os pares de eventos pessoais (ou clusters). O cálculo da similaridade entre eventos é realizado através de uma função. Esta função utiliza as informações de contexto do conjunto de fotografias capturadas em um evento. A função de similaridade é definida como na equação 2.2:

$$\varphi(e1, e2) = SS * SSW + TS * TSW + PS * PSW \quad (2.2)$$

Onde as funções SS, TS e PS medem o grau de similaridade espacial, temporal e de pessoas, respectivamente, entre os eventos e1 e e2. As variáveis SSW, TSW e PSW são os pesos dos estimadores de similaridade espacial, temporal e de pessoas, respectivamente.

As equações 3, 4 e 5 demonstram como o cálculo de similaridade espacial, temporal e de pessoas é realizado. Estas funções refletem a teoria da probabilidade no que concerne à união entre dois conjuntos, além de seguir o raciocínio que rege o coeficiente de Jaccard (1901), o qual é utilizado para comparar a similaridade e a diversidade entre conjuntos, sendo aplicável ao contexto de similaridade de eventos.

O cálculo da similaridade espacial entre eventos é feito com base na porcentagem da área de interseção das áreas dos dois eventos dividido pela união das áreas dos dois eventos (equação 2.3). Para realizar este cálculo, primeiramente é verificada a existência de alguma interseção entre as áreas dos dois eventos. Caso haja, o cálculo da similaridade é feito com base nessa interseção e, posteriormente, a área da geometria de cada evento é calculada. A função de similaridade espacial (SS) é definida como segue:

$$SS = \frac{\text{área}(\text{interseção}(e1, e2))}{(\text{área}(e1) + \text{área}(e2) - \text{área}(\text{interseção}(e1, e2)))} \quad (2.3)$$

O cálculo da similaridade temporal entre eventos é feito com base na porcentagem do intervalo em comum dos dois eventos dividido pela união do tempo total destes (equação 2.4). Para realizar este cálculo, são obtidos o tempo inicial e final dos dois eventos a partir das informações temporais das fotos mais antigas e mais recentes de cada um deles. Considerando que o evento pode ter iniciado antes de a primeira fotografia ser capturada, e terminado depois de a última fotografia ser capturada, um

intervalo de tempo é adicionado aos tempos inicial e final dos dois eventos. A função de similaridade temporal (TS) é definida como segue:

$$TS = \frac{\text{intervalo}(it_{se}, ft_{se})}{(\text{intervalo}(it_{e1}, ft_{e1}) + \text{intervalo}(it_{e2}, ft_{e2}) - \text{intervalo}(it_{se}, ft_{se}))} \quad (2.4)$$

Onde it_{se} e ft_{se} são os tempos inicial e final da interseção entre os eventos $e1$ e $e2$. it_{e1} , ft_{e1} , it_{e2} e ft_{e2} são, respectivamente, os tempos inicial e final do evento $e1$ e os tempos inicial e final do evento $e2$.

O cálculo da similaridade de pessoas em eventos é feito com base na porcentagem de pessoas em comum entre dois eventos dividido pela união do total de pessoas dos eventos (equação 2.5). Para realizar este cálculo, são obtidas as quantidades de pessoas distintas anotadas nos eventos $e1$ e $e2$ e a quantidade de pessoas em comum anotadas nas fotografias de ambos os eventos, utilizando o identificador (id) de cada pessoa cadastrada no banco de dados para realizar este cálculo. A função de similaridade de pessoas (PS) é definida como segue:

$$PS = \frac{|P_{e1} \cap P_{e2}|}{(|P_{e1}| + |P_{e2}| - |P_{e1} \cap P_{e2}|)} \quad (2.5)$$

Onde P_{e1} e P_{e2} são, respectivamente, os identificadores das pessoas presentes nos eventos $e1$ e $e2$.

2.2.3 Anotação semiautomática de pessoas em fotografias

Na anotação semiautomática de pessoas, foi utilizado o algoritmo proposto por Figueirêdo et al. (2015). Este algoritmo se baseia no uso de estimadores, que são estatísticas baseadas em uma ou mais informações contextuais (ou de conteúdo) das fotografias capturadas. A seguir são descritos os estimadores de contexto e de conteúdo utilizados nesta pesquisa, além da geração de lista de recomendações e da ponderação dos estimadores - passos subsequentes à utilização dos estimadores na anotação semiautomática de pessoas em fotografias digitais.

2.2.3.1 Estimadores baseados em contexto

A definição para os estimadores Recorrência Espacial e Recorrência Temporal utiliza a equação 2.6, para cada contato $c \in C$.

$$\text{Recorrência}(c, X) = \sum_{p \in X} \frac{\text{presente}(c, p)}{|X|}, \text{ presente}(c, p) = \begin{cases} 1 & \text{se } c \in C_p, \\ 0 & \text{caso contrário} \end{cases} \quad (2.6)$$

Onde $X \subseteq P$ é um conjunto de fotografias que depende da especificação do estimador e C_p é conjunto de contatos presentes na fotografia p .

|SR| Recorrência Espacial: indica a porcentagem de vezes que um contato foi anotado em uma fotografia dentro de um raio de d metros da localização geográfica da fotografia p_0 a ser anotada. Para calcular este estimador, tem-se que $X = \{p \mid \text{distância}(p_0, p) \leq d\}$ na equação 2.6.

|TR| Recorrência Temporal: indica a porcentagem de vezes que um contato foi anotado num intervalo t do momento em que a fotografia p_0 a ser anotada foi capturada. Para o cálculo deste estimador, considera-se $X = \{p \mid \text{tempo}(p_0) - t \leq \text{tempo}(p) \leq \text{tempo}(p_0) + t\}$ na equação 2.6.

|CO| Coocorrência: indica a porcentagem de vezes que um contato está presente em uma fotografia simultaneamente com outro contato que já foi anotado naquela fotografia. A coocorrência é calculada seguindo a definição de probabilidade condicionada, proposta por Kolmogorov (1933). Considerando dois contatos c_1 e c_2 , sendo $c_2 \in C_p$ (previamente anotado na fotografia p), a coocorrência de c_1 e c_2 é calculada pelas equações 2.7 e 2.8:

$$CO(c_1|c_2) = \frac{\sum_{p \in P} \text{coexistente}(c_1, c_2, p)}{\sum_{p \in P} \text{presente}(c_2, p)} \quad (2.7)$$

$$\text{coexistente}(c_1, c_2, p) = \begin{cases} 1 & \text{se } c_1, c_2 \in C_p, \\ 0 & \text{caso contrário} \end{cases} \quad (2.8)$$

Onde C_p e $\text{presente}(c, p)$ são definidos como na equação 2.6.

Tendo em vista que apenas um valor para o estimador CO pode ser utilizado para cada contato $c_1 \in C_p$, CO é computado para c_1 como definido na equação 2.9, considerando apenas o contato c_2 que tiver maior coocorrência com c_1 na coleção de fotos:

$$CO(c_1) = \text{Max}_{c_2 \in C_p} (CO_{c_1, c_2}) \quad (2.9)$$

2.2.3.2 Estimadores baseados em conteúdo

Além dos estimadores que se baseiam no contexto das fotografias, existem os estimadores que são obtidos a partir da análise do conteúdo delas. Nesta proposta, foi utilizada apenas um estimador baseado em conteúdo que se baseia em algoritmos de detecção e reconhecimento de faces.

O reconhecimento de faces utiliza o algoritmo de EigenFaces (TURK; PENLAND, 1991). Nesta etapa, uma pontuação para cada contato $c \in K$ (conjunto de

peças identificadas pelo sistema) é calculada e armazenada na base de dados. A etapa de reconhecimento facial retorna uma lista H_p com as n pessoas melhor classificadas em uma fotografia p ; ou seja, as n pessoas cuja face mais se assemelha à face detectada. Se uma pessoa é marcada em mais de uma face para a mesma fotografia, a face na qual a pessoa obteve a maior pontuação prevalecerá. Finalmente, o usuário é então convidado a confirmar as pessoas listadas em H_p .

O estimador baseado em conteúdo utiliza a pontuação da etapa de reconhecimento facial e é definido como segue:

[FR] Reconhecimento de Faces: maior pontuação do contato entre os valores retornados pelo algoritmo de reconhecimento de faces aplicado às faces detectadas na fotografia.

2.2.3.3 Geração da lista de recomendações

De modo a auxiliar no processo de anotação de pessoas presentes nas fotografias, a lista de recomendações é gerada utilizando vários estimadores para avaliar a probabilidade $Pr_{c,p}$ de uma pessoa $c \in P$ estar em C_p . Esta probabilidade é calculada utilizando uma soma ponderada dos estimadores previamente discutidos, conforme a equação 2.10:

$$Pr_{c,p} = \sum_{e \in E} estimador(e, c, p) * peso(e) \quad (2.10)$$

Onde E é o conjunto de todos os estimadores, *estimador* é uma função que calcula o valor do estimador e para o contato c na fotografia p , e *peso* é uma função que retorna o peso do estimador e .

Após a combinação dos estimadores, uma classificação ordenada de forma descendente é gerada a partir do valor de $Pr_{c,p}$. A próxima subseção trata das ponderações dos estimadores.

2.2.3.4 Ponderação dos estimadores

Existem algumas técnicas de ponderação que utilizam heurísticas para determinar quais estimadores devem possuir maior peso na combinação do que outros. No entanto, percebe-se que tais técnicas não possuem uma justificativa matemática para provar que as ponderações adotadas são as ideais.

Como sugerido por Figueirêdo et al. (2015), nesta pesquisa foi utilizada uma função de inclinação para ponderação dos estimadores, que é a taxa de variação na linha de regressão entre dois pontos. De acordo com os autores, esta função foi adotada após a realização de experimentos com alguns métodos de ponderações

encontrados na literatura. Para a filtragem dos estimadores, foi utilizada uma regressão linear de multivariáveis usando o método de mínimos quadrados para a remoção dos estimadores do somatório ponderado.

Os estimadores utilizam as informações de contexto extraídas de uma fotografia e as analisam juntamente com as informações das fotografias já anotadas pelo sistema, estimando a probabilidade de cada um dos contatos do usuário estar presentes na fotografia.

Uma combinação dos resultados dos estimadores é realizada após cada um deles terem calculado o valor da estimativa de cada contato estar presente na fotografia. Então, esta combinação gera um valor único para cada contato, representando a probabilidade de o contato estar presente na fotografia. Por fim, uma lista com os contatos mais propensos a estarem na fotografia é gerada e apresentada ao usuário de modo a auxiliá-lo na anotação das pessoas presentes na fotografia.

2.2.3.4.1 Regressão linear múltipla

Os métodos de regressão tornaram-se um componente integral de qualquer análise de dados envolvida na descrição da relação entre uma variável de resposta e uma ou mais variáveis explicativas. Muitas vezes a variável de resultado é discreta, assumindo dois ou mais valores possíveis. O modelo de regressão logística é o modelo de regressão mais utilizado para a análise desses dados (HOSMER; LEMESHOW, 2003).

Os fatores para incluir uma variável em um modelo de regressão linear múltipla podem variar de um problema para o próximo e de uma disciplina científica para outra. A abordagem tradicional na construção de modelos estatísticos é encontrar o modelo mais parcimonioso que explica os dados. Quanto mais variáveis no modelo, maior se torna a estimativa do erro e mais dependente o modelo fica dos dados observados.

A justificativa para minimizar o número de variáveis no modelo é que o modelo resultante é mais provável ser numericamente estável, e é mais facilmente adotado para uso. Quanto mais variáveis forem incluídas num modelo, maiores serão os erros padrões estimados e mais dependentes serão os modelos nos dados observados (HOSMER; LEMESHOW, 2003).

A regressão linear múltipla relaciona 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 2.11:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \varepsilon \quad (2.11)$$

2.2.4 Detecção de Faces

Inicialmente, foi utilizado o algoritmo de classificação proposto por Viola e Jones (2001) e aperfeiçoado por Lienhart e Maydt (2002) para detecção de faces. De acordo com Andrade et al. (2016), a maior parte dos trabalhos na área de anotação de pessoas utiliza este algoritmo para detectar as faces das pessoas presentes nas fotografias. Dentre os trabalhos pesquisados pelos autores, 61% deles utilizou o algoritmo proposto por Viola e Jones (2001).

No entanto, durante alguns experimentos realizados em se tratando de anotação automática de pessoas, constatou-se que o algoritmo utilizado para detecção de faces não apresentou uma boa taxa de acerto. As faces eram detectadas apenas quando a foto estava bem iluminada e as faces estavam na posição total frontal, sem nenhum tipo de variação de pose. Além disso, houve casos em que o algoritmo indicou a existência de faces quando não havia nenhuma.

A figura 4 apresenta exemplos de faces detectadas (ou não) utilizando o algoritmo proposto por Viola e Jones (2001). Na figura 4 (a) vê-se que, apesar de duas pessoas estarem presentes na fotografia, nenhuma face foi detectada. O algoritmo identificou uma face onde na realidade existia um violão. Na figura 4 (b) pode-se ver que três pessoas estão presentes na fotografias, mas o algoritmo apenas detectou a face que estava mais próxima à câmera. Já na figura 4 (c), percebe-se que as duas pessoas presentes na fotografias estão na mesma distância em relação à câmera, no entanto a face de apenas uma delas foi detectada; a outra não foi detectada devido a uma oclusão parcial da face (o microfone cobriu a boca e o queixo da pessoa). Por fim, na figura 4 (d) pode-se perceber que quinze pessoas estão presentes na fotografia (a qual foi obtida da base disponibilizada por Gallagher e Chen (2008)), porém o algoritmo não conseguiu detectar nenhuma face. Isso se deu devido à distância que as faces estavam em relação à câmera.

Figura 4 – Exemplos de faces detectadas utilizando o algoritmo de Viola e Jones (2001)



Fonte: Elaborado pelo autor

Diante disso, optou-se por utilizar um algoritmo que tivesse uma taxa de acerto superior (de acordo com o Face Detection Data Set and Benchmark¹) ao algoritmo proposto por Viola e Jones. Após a tentativa de se executar algum dos algoritmos disponíveis no Benchmark citado, foi possível instalar o algoritmo proposto por Markus et al. (2013), o qual apresentou uma taxa de acerto 20% superior ao algoritmo proposto por Viola e Jones (2001) com os dados do Benchmark citado.

Na figura 5, pode-se ver exemplos de faces detectadas utilizando o algoritmo proposto por Markus et al. (2013). Ao comparar as faces detectadas nesta figura com as que foram detectadas utilizando o algoritmo de Viola e Jones (figura 4), vê-se que houve a detecção de maior número de faces. Na figura 5 (d), quando comparada com figura 4 (d) - na qual nenhuma face foi detectada - percebe-se que o algoritmo de Markus et al. é mais robusto, sendo capaz de detectar faces não tão próximas à lente do dispositivo fotográfico. Além disso, na figura 5 (a), pode-se observar que, mesmo havendo uma oclusão parcial na face do homem de camisa verde (com o microfone cobrindo sua boca e nariz), sua face foi detectada corretamente pelo algoritmo.

¹ <http://vis-www.cs.umass.edu/fddb/results.html>

Figura 5 – Exemplos de faces detectadas utilizando o algoritmo proposto por Markus et al. (2013)



Fonte: Elaborado pelo autor

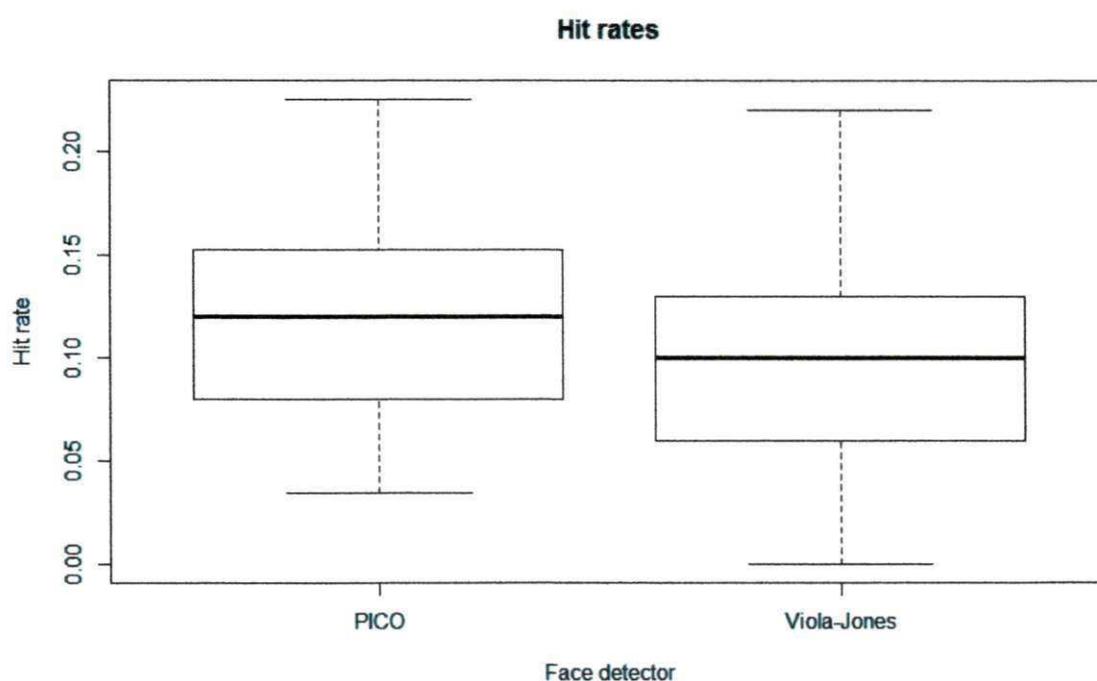
Para verificar o impacto dos dois detectores de faces citados no contexto de anotação automática e semiautomática de pessoas, foram realizados alguns experimentos. O propósito de se utilizarem ambas as soluções é verificar se o uso da solução PICO traz algum incremento na taxa de acerto em anotação automática e semiautomática de pessoas em fotografias, quando comparado com o uso da solução Viola-Jones.

Na figura 6, pode-se ver os resultados da execução do algoritmo de anotação automática de pessoas utilizando os dois detectores de faces mencionados nesta pesquisa. É possível visualizar que os resultados foram próximos, havendo sobreposição nos boxplots porém, sem sobreposição de medianas - o que indica que é muito provável que os resultados utilizando o detector PICO sejam superiores aos da utilização do detector Viola-Jones. Assim, para os demais experimentos, foi utilizada a solução PICO para detecção de faces.

Apesar de o resultado do detector de faces PICO ter sido 20% superior ao Viola-Jones no benchmark FDDB, é importante ressaltar que nesta pesquisa não foram avaliados os desempenhos dos detectores de forma isolada; porém, foi avaliado o desempenho dos detectores na anotação de pessoas. Na anotação automática, além

da etapa de detecção de faces, também há a etapa reconhecimento. Além disso, no FDDB a métrica utilizada para avaliar o desempenho dos detectores foi a curva ROC, enquanto que nesta pesquisa foi utilizada a métrica taxa de acerto.

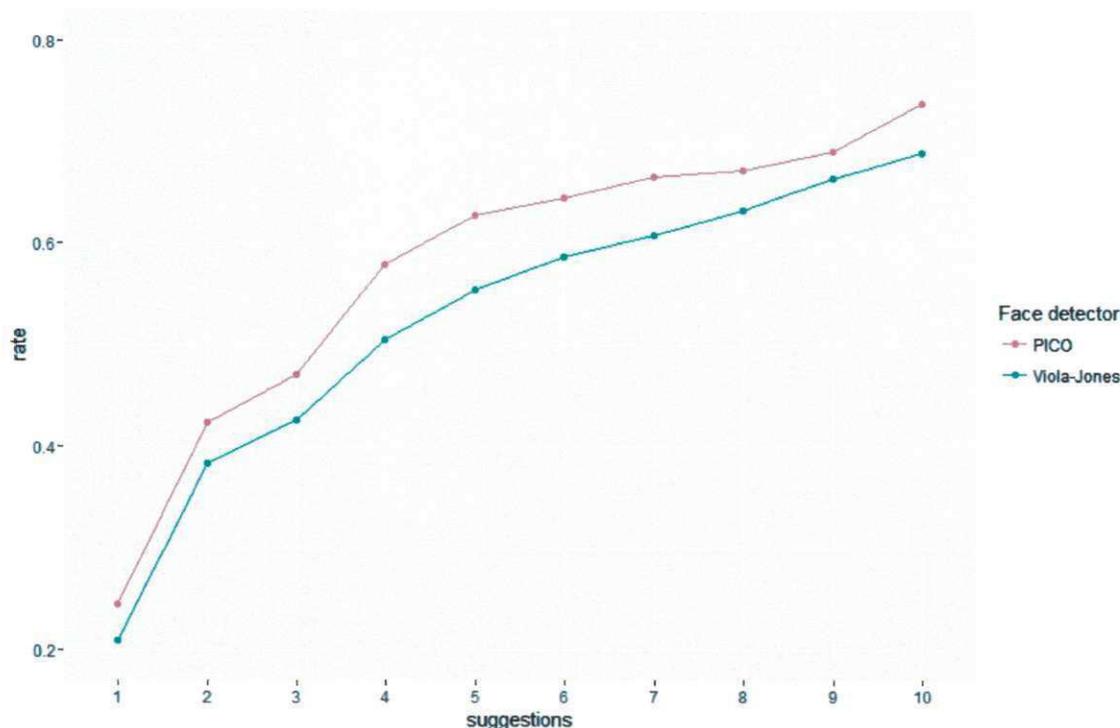
Figura 6 – Comparação de detectores de face utilizados na anotação automática de pessoas



Fonte: Elaborado pelo autor

Da mesma forma que foi realizada uma comparação entre as taxas de acerto para a anotação automática de pessoas em fotografias utilizando as soluções para detecção de faces Viola-Jones e PICO, foi feito o mesmo para anotação semiautomática. Na figura 7, pode-se visualizar uma comparação entre o uso dos detectores de face citados na anotação semiautomática de pessoas.

Figura 7 – Comparação de detectores de face utilizados na anotação semiautomática de pessoas



Fonte: Elaborado pelo autor

A figura 7 apresenta os resultados da execução do algoritmo para a coleção de fotografias utilizada. Para cada evento na coleção de fotografias, o algoritmo foi executado utilizando o detector de faces Viola-Jones; e, em seguida, o algoritmo foi executado utilizando o detector PICO. Como se pode notar, o uso do detector de faces proposto por Markus et al. (2013) - PICO - trouxe um incremento na taxa de acerto em se tratando de anotação semiautomática de pessoas para a coleção de fotos utilizada. Por isso, optou-se por utilizar este detector nos demais experimentos.

2.3 Considerações Finais

Neste capítulo, apresentou-se a fundamentação teórica deste trabalho. Foram expostos os principais conceitos relacionados com anotação em fotografias digitais, focando nos princípios básicos sobre detecção de eventos pessoais, com ênfase no funcionamento do algoritmo de agrupamento DBSCAN; e nos estimadores utilizados na anotação semiautomática de pessoas.

Os algoritmos estudados e utilizados foram elencados com base na literatura. No próximo capítulo, são apresentados alguns trabalhos relacionados à pesquisa realizada.

3 Trabalhos Relacionados

Esta seção apresenta um levantamento bibliográfico das pesquisas relacionadas à recuperação e anotação de pessoas e eventos em fotografias digitais.

3.1 Anotações de pessoas

De acordo com a literatura pesquisada, verificou-se que os trabalhos na área de anotações de pessoas em fotografias digitais utilizam informações de conteúdo, contexto ou ambos para anotar as pessoas presentes nas fotos. Informações de contexto incluem data e hora de captura, localização geográfica e tags que podem ser associadas às fotografias. Já as informações de conteúdo incluem aspectos que estão presentes nas fotografias, como faces detectadas e roupas das pessoas presentes.

Naaman et al. (2005) utilizaram a ideia de que há padrões para o aparecimento das pessoas nas fotografias, viabilizando a anotação de pessoas de forma semiautomática. Os padrões idealizados pelos autores são utilizados como estimadores para calcular a chance de uma determinada pessoa estar numa fotografia. O sistema sugere uma lista de pessoas com alta probabilidade de estarem numa dada fotografia por meio do cálculo destes estimadores. A probabilidade de uma determinada pessoa estar numa fotografia é calculada a partir da soma de todos os estimadores, utilizando o mesmo peso para cada estimador. Dessa forma, não há ponderação dos estimadores de acordo com as coleções de fotografias. Um exemplo de estimador utilizado neste trabalho foi a recorrência espacial.

A metodologia utilizada por Jamil e Sa'dan (2014) trata de anotação automática de pessoas realizando reconhecimento de faces e características adicionais, como cor da roupa e data da foto. As autoras utilizaram três métricas de similaridade baseadas em distância (Euclidiana, de Manhattan e de Canberra). Ao final, todas as combinações de métricas e características utilizadas (cor da roupa, data da foto) foram avaliadas. Foi visto que os melhores resultados foram obtidos ao utilizar as três características estudadas (face, cor da roupa e data) e a distância Euclidiana como métrica de similaridade (obtendo uma taxa de acerto de 82%).

A ferramenta proposta por O'Hare e Smeaton (2009) oferece navegação, busca e anotação semiautomática de fotografias utilizando análises de conteúdo de imagem e contexto no qual a fotografia foi capturada. A principal anotação semiautomática realizada pelo sistema proposto é o da identidade das pessoas nas fotografias, mediante a análise do conteúdo das imagens e do contexto associado. Para a elaboração da lista de pessoas com maior probabilidade de estarem nas fotografias, foram utilizadas as

técnicas de somatório e somatório ponderado para a combinação de estimativas da presença de uma pessoa em uma fotografia, as quais utilizam informações de contexto e conteúdo.

Choi et al. (2011) criaram um framework para a fusão dos resultados de várias ferramentas de reconhecimento de faces existentes nas redes sociais para a execução de uma anotação automática e colaborativa de pessoas presentes nas fotografias. Uma regra de decisão Bayesiana foi utilizada para realizar a fusão dos resultados das ferramentas usadas.

Yagnik e Islam (2007) desenvolveram uma metodologia para reconhecer pessoas em uma imagem, sem levar em consideração aspectos como pose, idade e expressão das pessoas mesmo sem anotações prévias. O método desenvolvido é chamado de aprendizagem consistente e utiliza técnicas de aprendizagem de máquina para extrair a coocorrência de texto-imagem e aprende um conjunto de modelos de face consistentes a partir de páginas da Internet.

A abordagem de Madhumathi e Thanamani (2014) trata de anotação de faces baseada em pesquisas. A ideia básica é obter imagens faciais com nomes associados que são adquiridas pela Internet. Em seguida, realiza-se a extração e a indexação de aspectos faciais, diferenciando as faces nas quais uma pessoa está sorrindo das faces nas quais uma pessoa está séria, por exemplo. Por fim, faces semelhantes são agrupadas e é feita a anotação de acordo com um algoritmo de votação.

Brenner, Mirza e Izquierdo (2014) desenvolveram uma abordagem que envolve reconhecimento de faces, mas não necessita de um modelo classificação treinado. A ideia geral é reconhecer todas as faces dentro da base de dados de forma simultânea e utilizar o parecer de usuários na internet para melhorar a acurácia da detecção. A obtenção das opiniões de usuários foi feita através de um jogo disponibilizado na internet, no qual em vários momentos os usuários precisam informar se duas faces da base de dados são iguais ou não.

Figueirêdo et al. (2015) utilizaram as informações de contexto (data, hora, localização geográfica) e de conteúdo (detecção e reconhecimento de faces) das fotografias para realizarem as anotações e concluíram que a taxa de acerto da técnica proposta (utilizando informações de contexto e conteúdo conjuntamente) foi superior a técnicas que utilizam apenas informações de conteúdo ou contexto isoladamente.

Zhang, Zhang e Tretter (2010) elaboraram uma técnica para agrupamento de pessoas em fotografias pelas roupas para uma posterior utilização em algoritmos de reconhecimento de faces. Neste trabalho não foi proposta uma solução para anotação automática de pessoas presentes nas fotografias.

Gallagher et al. (2008) utilizaram as informações de agendas pessoais como

contexto para anotar as fotografias com as informações do evento. Em outros trabalhos, utilizaram as roupas das pessoas (GALLAGHER; CHEN, 2008), o tempo (GALLAGHER; CHEN, 2009) e os relacionamentos sociais (WANG et al., 2010) para melhorar o desempenho de reconhecimento de faces. Para algumas coleções de fotografias testadas, a acurácia subiu de 80% para 86%.

Cooray e O'Connor (COORAY; O'CONNOR, 2009; COORAY, 2008; COORAY et al., 2006; SADLIER et al., 2008) utilizaram a detecção do evento de forma conjunta com o reconhecimento de faces. No trabalho, a detecção de eventos não focou na utilização de informações de contexto (ex: localização geográfica; data e hora da fotografia); ao invés disso, a solução se debruçou na utilização de informação de conteúdo das fotografias, como as faces detectadas e as roupas que as pessoas estavam utilizando.

Kim, El-Saddik e Jung (2012) propuseram uma aplicação (VizFaceCo) que utiliza reconhecimento de faces para anotar semiautomaticamente as pessoas que estão presentes em fotografias. Além disso, este sistema utiliza um modelo (FCON) que se baseia em uma rede de coocorrência de pessoas que estão presentes nas fotos para recomendar amigos. No entanto, VizFaceCo não utiliza as informações de contexto das fotografias para a anotação semiautomática de pessoas.

Lin et al. (2010) propuseram um modelo probabilístico que combina os domínios de localização, pessoas e eventos para a anotação de informações sobre esses domínios em fotografias, utilizando o reconhecimento de faces para a anotação de pessoas e as informações de contexto para detecção de eventos. As informações de contexto são usadas apenas para melhorar o desempenho do reconhecimento de faces.

Hulsebosch e Ebben (2008) usaram informações sobre a localização das pessoas para melhorar o desempenho do reconhecimento de faces. As localizações de pessoas são descobertas por meio de Bluetooth e RFID. Esta solução não gera uma lista de sugestões de pessoas que podem estar presentes em fotografias usando anotação semiautomática, mas considera apenas as faces que são detectadas automaticamente.

O sistema MMM (DAVIS et al., 2004; DAVIS et al., 2005; DAVIS et al., 2006) visa melhorar a anotação automática de pessoas usando análise de dados para combinar informações de contexto. No entanto, o sistema não usa informações de contexto e de conteúdo para gerar anotações semiautomáticas de pessoas.

PhotoGeo (LACERDA et al., 2008a; LACERDA et al., 2008b) é uma ferramenta que visa auxiliar na anotação, busca e armazenamento de pessoas e eventos. Ele também usa uma soma de estimadores para sugerir anotação semiautomática, comprometendo assim a taxa de acerto.

O sistema ACRONYM (MONAGHAN; O'SULLIVAN, 2007) apresenta uma abor-

dagem para anotar pessoas e lugares de forma semiautomática. O sistema usa dados de GPS, *Web semantics*, *Web services* e redes sociais para gerar sugestões. A ferramenta também usa a correlação de Pearson para gerar pesos para os estimadores. No entanto, ACRONYM não permite a adição de novos estimadores.

3.2 Anotações de eventos

No tocante à organização de fotografias em eventos, as pesquisas na área seguem abordagens distintas. Alguns trabalhos procuram por anotar o evento na fotografia, enquanto outros buscam agrupar as fotografias do mesmo evento, independentemente da anotação dos eventos. Essas abordagens de agrupamento e anotação podem englobar três tipos de eventos: eventos pessoais, sociais e colaborativos (BRENNER; IZQUIERDO, 2012; ZIGKOLIS et al., 2014). Os eventos pessoais são os que referenciam um acontecimento pessoal, como “aniversário de tia Rosa”. Os eventos sociais são aqueles que representam um acontecimento na sociedade, como, por exemplo, o tradicional desfile que ocorre no dia 7 de setembro. Já os eventos colaborativos incluem tanto os eventos pessoais quanto os sociais, porém suas informações são fornecidas pela colaboração dos usuários de um sistema. Essa colaboração se dá quando os usuários cadastram as informações dos eventos (nome, data, local) no sistema.

Andrade et al. (2016) realizaram uma pesquisa envolvendo anotações de pessoas e eventos em fotografias digitais. Neste trabalho, os autores se dispuseram a fazer uma pesquisa extensa nestas áreas. Foram apresentadas taxonomias na área no que concerne ao tipo de anotação (de pessoas, de eventos, de localização ou de outras informações de contexto), ao processo de anotação (automático, semiautomático ou manual), à fonte utilizada para realizar as anotações (conteúdo da imagem, informações de contexto associadas à imagem e redes sociais) e à forma de representação das anotações (ontologias, banco de dados relacionais ou baseada em RDF). Além disso, vários trabalhos foram comparados na área de anotação de pessoas e na área de anotação de eventos em fotografias, sob diversos aspectos, como por exemplo, se o trabalho detecta eventos não conhecidos previamente.

A ferramenta proposta por Zigkolis et al. (2014) - CrEve - é um framework de anotação colaborativa de eventos. Esse framework extrai informações de sites de mídias sociais com o objetivo de facilitar a anotação de grandes coleções de mídias com informações de eventos. Os autores comentam sobre a necessidade de se organizar e indexar a grande quantidade de conteúdo gerada em sites de mídia social de forma a tornar os mecanismos de busca eficientes. Nos experimentos, mostram-se que as consultas textuais possuem uma grande relevância para a recuperação de fotografias de um mesmo evento. Além disso, eles mencionaram a importância dos aspectos espaciais e temporais em consultas compostas.

Lim, Tian e Mulhem (2003) apresentam um modelo de detecção de eventos com base em processamento de imagem. Neste trabalho, as imagens são comparadas e, de acordo com a similaridade, relaciona-as ao respectivo evento. O modelo proposto apresenta bons resultados, porém os eventos são detectados como, por exemplo, piscina, parque. Esses eventos abstratos são previamente definidos pelo usuário, deste modo um novo evento nunca é criado.

Brenner e Izquierdo (2012) apresentam um framework para detectar eventos sociais e recuperar fotografias em coleções colaborativas de fotografias digitais. Os autores propuseram um framework que trata o problema da relação entre fotografias e eventos de maneira completa, fazendo a detecção e a recuperação de fotografias envolvidas nos eventos. É importante ressaltar que o trabalho depende de processamento digital de imagens e trabalha 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. O trabalho dos autores apresentou bons resultados, porém os eventos obtidos não correspondem aos eventos reais.

Mezaris et al. (2014) apresentaram uma discussão sobre os eventos da vida real em arquivos multimídia. Nove estudos foram destacados e estudados em termos de fenômenos naturais e detecção de eventos em vídeos e coleções de fotos. O trabalho dos autores mostra a presença de detecção de eventos em várias áreas e destaca a importância do contexto em fotografias e vídeos. No entanto, os autores não propõem nenhuma abordagem para detecção de eventos sociais.

Feng et al. (2014) abordaram a presença de eventos em redes sociais de forma diferente. Ao invés de realizar buscas entre os eventos nas redes sociais, este trabalho busca por pessoas influentes que possam ser organizadoras de eventos. Apesar do fato deste trabalho não contemplar detecção de eventos, são mostradas várias soluções para encontrar pessoas influentes. Assim sendo, eventos poderiam ser encontrados através destas pessoas.

A proposta de Rabbath, Sandhaus e Boll (2012) é voltada para aplicações de fotografias baseadas na rede social *Facebook*. O trabalho lida com os metadados fornecidos pelo *Facebook* e apresenta uma solução para detectar fotografias que pertencem ao mesmo evento e que estão distribuídas entre vários amigos. Os autores 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 citada.

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.

Presti e Cascia (2014) elaboraram um framework que organiza múltiplas sequências de fotografias concorrentes baseado na proximidade temporal e similaridade do conteúdo visual. Apesar de organizar as fotos que pertencem a uma mesma janela espaço-temporal, o framework não viabiliza a detecção de eventos compartilhados.

Psallidas et al. (2013) abordam a identificação de eventos não conhecidos em redes sociais, propondo uma nova técnica para este fim. Essa técnica utiliza informações associadas às publicações coletadas (usuário que publicou, data/hora, texto descritivo) nas redes sociais. Com essas informações, funções de similaridade temporal, espacial e textual são utilizadas 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. Uma das limitações deste trabalho é a dependência das redes sociais, fazendo com que a detecção se dê somente com eventos que são discutidos nas redes.

Bouselmi, Fohr e Illina (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 a um evento. Dessa forma, o processo de detecção de eventos torna-se binário e a solução proposta não identifica possíveis eventos específicos. A solução proposta no trabalho apresenta bons resultados, contudo não identifica os eventos aos quais as publicações estão relacionadas.

A ferramenta proposta por Gao et al. (2013) - GESODECK - detecta e acompanha eventos sociais em redes sociais, como o Twitter. O propósito da ferramenta é identificar o comportamento de usuários na rede social, usando as informações geográficas do post para apontar a alternância de usuários que publicam conteúdo de uma área geográfica específica. Os autores definem que essa alternância de usuários ultrapassa o padrão natural dessas regiões quando, em um determinado período de tempo, o número de usuários alternando publicações na região está maior que o normal da região.

Jang, Yoon e Cho (2009a) foram os primeiros a utilizarem o fato da existência de várias câmeras em um mesmo evento. Os estudos anteriores em fotografias digitais focavam em bases massivas de fotografias não relacionadas ou coleções privadas de usuários. Jang, Yoon e Cho (2009b) propuseram o agrupamento de fotografias que foram capturadas por um grupo de pessoas durante viagens utilizando informações

espaço-temporais, sendo o agrupamento segundo as preferências do dono de cada câmera.

Paniagua et al. (2013) propuseram uma nova abordagem para detecção de eventos. 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 a semelhança entre os eventos pessoais de usuários distintos no que concerne a data e horário em que as fotografias foram capturadas e a função de similaridade espacial, no que concerne a localização das fotografias. De certo modo, a proposta dos autores leva em consideração o conceito de evento compartilhado, pois para que um evento social seja detectado, é necessário que haja a interseção entre os eventos pessoais dos usuários.

Além destes exemplos citados da literatura, a Google dispõe de uma ferramenta de organização de fotografias - Google Photos¹ -, a qual permite a criação de álbuns compartilhados entre diferentes usuários. Para cada álbum compartilhado, todos os usuários participantes podem adicionar fotografias. Deste modo, pode-se dizer que um álbum compartilhado corresponde a um evento compartilhado. No entanto, é importante ressaltar que o Google Photos¹ não detecta eventos compartilhados de forma automática.

Figueirêdo et al. (2012) trataram sobre detecção de eventos compartilhados no contexto de coleções de fotografias pessoais. Para tal, são utilizadas redes sociais de compartilhamentos de fotografias para a descoberta de eventos compartilhados na lista de contatos de um usuário. Informações espaciais e temporais das fotografias são utilizadas para auxiliar no processo. A proposta dispensa o processamento de imagens para detectar eventos, mas requer muitos metadados, o que pode torná-la dependente a eles.

Sansone et al. (2017) propuseram uma abordagem para realizar a sincronização de galerias de fotos de um mesmo evento. Eles utilizaram a informação temporal das fotografias, além de redes neurais para identificar semelhanças entre elas. Apesar do foco do trabalho ser em eventos esportivos, é realizado o agrupamento de fotografias em eventos, utilizando assim o conceito de evento compartilhado. Porém, é importante ressaltar que este trabalho não trata de anotações de pessoas.

¹ photos.google.com

3.3 Comparação dos trabalhos relacionados

Nesta seção, é apresentada uma comparação entre os trabalhos estudados que tratam da anotação de pessoas ou eventos em coleções de fotografias digitais.

É importante ressaltar que o trabalho proposto baseia-se na ideia de evento compartilhado de Figueirêdo et al. (2012), utilizando-a para melhorar o desempenho da anotação automática e semiautomática de pessoas em fotografias (FIGUEIRÊDO et al., 2015).

A tabela 1 apresenta um estudo comparativo analisando as seguintes características:

C1: Utilização de anotação automática de pessoas: quando o trabalho anota as pessoas de forma automática ou apenas utiliza estas anotações com alguma finalidade, como por exemplo, organizar as coleções de fotografias em eventos.

C2: Utilização de anotação semiautomática de pessoas: quando o trabalho anota as pessoas de forma semiautomática ou apenas utiliza estas anotações.

C3: Utilização do conceito de evento: quando o trabalho realiza a organização das fotografias em eventos. Um trabalho organiza as fotos em eventos quando um usuário pode explicitamente atribuir um evento às fotografias, ou quando o trabalho cria eventos para organizar as coleções de fotografias.

C4: Utilização da ideia de evento compartilhado: quando o trabalho lida com a ideia de vários usuários estarem presentes em um mesmo evento, mas não realiza a detecção de um evento compartilhado.

C5: Detecção de eventos compartilhados: quando o trabalho organiza as fotografias em eventos e realiza a descoberta de eventos similares. Ou seja, além de realizar a organização de fotografias em eventos, o trabalho também viabiliza a descoberta de eventos compartilhados dentre os que foram detectados previamente.

Tabela 1 – Comparação dos trabalhos relacionados

Trabalho	C1	C2	C3	C4	C5

Trabalho	C1	C2	C3	C4	C5
(NAAMAN et al., 2005)	Não	Sim	Sim	Não	Não
(JAMIL; SA'DAN, 2014)	Sim	Não	Não	Não	Não
(O'HARE; SMEATON, 2009)	Não	Sim	Não	Não	Não
(YAGNIK; ISLAM, 2007)	Sim	Não	Não	Não	Não
(MADHUMATHI; THANAMANI, 2014)	Sim	Não	Não	Não	Não
(FIGUEIRÊDO et al., 2015)	Não	Sim	Sim	Não	Não
(CHOI et al., 2011)	Sim	Não	Não	Não	Não
(BRENNER; MIRZA; IZQUIERDO, 2014)	Não	Não	Não	Não	Não
(GALLAGHER et al., 2008)	Sim	Não	Sim	Não	Não
(DAVIS et al., 2004)	Sim	Não	Não	Não	Não
(KIM; EL-SADDIK; JUNG, 2012)	Não	Sim	Não	Não	Não
(LIN et al., 2010)	Não	Sim	Sim	Não	Não

Trabalho	C1	C2	C3	C4	C5
(COORAY; O'CONNOR, 2009)	Não	Sim	Sim	Não	Não
(HULSEBOSCH; EBBEN, 2008)	Sim	Não	Não	Não	Não
(LACERDA et al., 2008b)	Sim	Sim	Sim	Não	Não
(MONAGHAN; O'SULLIVAN, 2007)	Não	Sim	Não	Não	Não
(ZHANG; ZHANG; TRETTER, 2010)	Não	Não	Sim	Não	Não
(JANG; YOON; CHO, 2009b)	Não	Não	Sim	Sim	Não
(ZIGKOLIS et al., 2014)	Não	Não	Sim	Não	Não
(BRENNER; IZQUIERDO, 2012)	Não	Não	Sim	Não	Não
(PRESTI; CASCIA, 2014)	Não	Não	Sim	Não	Não
(LIM; TIAN; MULHEM, 2003)	Não	Não	Sim	Não	Não
(FENG et al., 2014)	Não	Não	Não	Não	Não
(RABBATH; SANDHAUS; BOLL, 2012)	Não	Não	Sim	Sim	Não

Trabalho	C1	C2	C3	C4	C5
(NAKAJI; YANAI, 2012)	Não	Não	Sim	Não	Não
(PSALLIDAS et al., 2013)	Não	Não	Sim	Não	Não
(BOUSELMI; FOHR; ILLINA, 2012)	Não	Não	Não	Não	Não
(GAO et al., 2013)	Não	Não	Sim	Não	Não
(PANIAGUA et al., 2013)	Não	Não	Sim	Sim	Não
Google Photos	Não	Sim	Sim	Sim	Não
(FIGUEIRÉDO et al., 2012)	Não	Não	Sim	Sim	Sim
Firmino, 2017	Sim	Sim	Sim	Sim	Sim

Fonte: Elaborado pelo autor

Os trabalhos de Mezaris et al. (2014) e Andrade et al. (2016) 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 nem de pessoas. Com o estudo apresentado na tabela 1, se evidencia o fato de que as anotações de pessoas e de eventos não são abordadas de maneira conjunta pelos trabalhos citados. Nenhum trabalho dentre a literatura pesquisada viabiliza a descoberta de eventos compartilhados em conjunto com a anotação de pessoas.

Como pode ser visto na tabela 1, apenas este trabalho aborda as cinco características. Gallagher et al. (2008) e Cooray e O'Connor (2009) utilizaram a anotação automática e semiautomática de pessoas em conjunto com o conceito de evento, sem levar em consideração a existência de múltiplas fotografias de diferentes usuários em um mesmo evento. É importante ressaltar que alguns trabalhos dentre a literatura pesquisada - Jang, Yoon e Cho (2009b), Rabbath, Sandhaus e Boll (2012), Google Photos e Paniagua et al. (2013) - lidaram com a ideia de evento compartilhado, porém não

realizando a detecção deste tipo de evento. Rabbath, Sandhaus e Boll (2012) apenas viabilizaram a descoberta de eventos compartilhados dentre os eventos pessoais de usuários, enquanto que Figueirêdo et al. (2012) realizaram a detecção de eventos pessoais e, em seguida, de eventos compartilhados.

3.4 Considerações Finais

Neste capítulo foram apresentados alguns trabalhos relacionados à pesquisa realizada nesta dissertação, incluindo estudos nas áreas de anotações de pessoas e eventos em coleções de fotografias digitais. Conforme apresentado na tabela 1, fica claro que os trabalhos estudados não abordam a anotação de pessoas e eventos compartilhados 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.

No próximo capítulo, está descrito em detalhes a abordagem de anotação de pessoas e eventos proposta neste trabalho.

4 Anotação de pessoas utilizando evento compartilhado

A informação das pessoas presentes pode ser utilizada para melhorar a organização e recuperação de coleções de fotografias. De acordo com Naaman et al. (2005), a identificação das pessoas presentes na cena é considerada a informação mais importante, permitindo que uma pessoa se recorde de uma fotografia. Atualmente, muitas câmeras e smartphones apresentam as faces detectadas nas fotografias capturadas. Alguns dispositivos equipados com o sistema Android anotam as pessoas presentes de forma automática, porém com precisão baixa, visto que se faz necessário que não haja oclusões na face, nem variações de pose.

A ideia de que diversas pessoas estiveram presentes em um determinado evento no qual cada pessoa capturou diversas fotografias com seu smartphone ou câmera digital é geralmente desconsiderada em se tratando de organizar fotografias utilizando eventos. A utilização do conceito de evento compartilhado pode ser útil para anotar pessoas em fotografias, considerando que a mesma pessoa pode estar em fotografias capturadas por diferentes dispositivos.

Tradicionalmente, quando se lida com várias fotos de um usuário, as pessoas são anotadas com base nos contatos do mesmo. No entanto, ao lidar com fotos de vários usuários, mesmo que eles tenham participado de um evento pessoal em comum, a anotação de pessoas é feita sem considerar este fato. A ideia deste trabalho é realizar a anotação de pessoas usando todos os participantes do mesmo evento compartilhado, permitindo a anotação e propagação de anotações de pessoas em fotos de diferentes usuários.

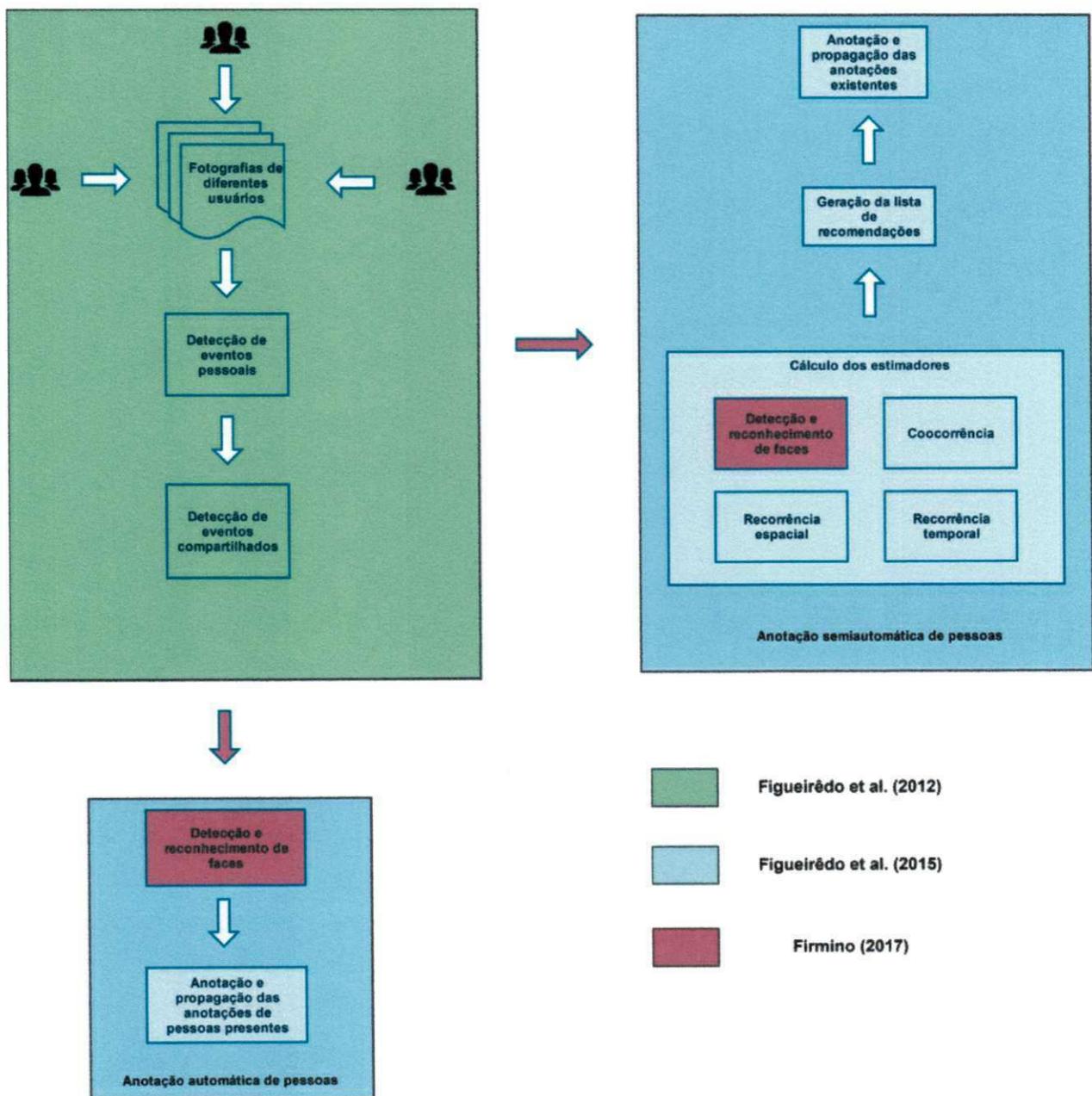
O processo de anotação de fotografias pode ser feito de três diferentes maneiras: manualmente, na qual o usuário realiza a anotação sem o auxílio de uma aplicação; semiautomaticamente, na qual uma aplicação ajuda o usuário a realizar a anotação; e automaticamente, onde uma aplicação anota as fotografias.

Desta forma, a solução proposta se baseia na realização de anotação automática e semiautomática de pessoas. O diferencial deste trabalho está no fato de se utilizar o conceito de evento compartilhado para realizar anotações de pessoas em fotografias. A solução proposta neste capítulo para anotação de pessoas utilizando evento compartilhado é dividida em três escopos distintos:

- **Detecção de eventos pessoais e compartilhados:** Separação das coleções de fotografias em eventos pessoais e, posteriormente, descoberta da existência de fotografias de um mesmo evento em coleções de fotografias de usuários distintos;

- **Anotação automática de pessoas em fotografias:** Anotação e propagação das anotações de pessoas utilizando detecção e reconhecimento de faces;
- **Anotação semiautomática de pessoas em fotografias:** Anotação de pessoas utilizando estimadores baseados nas informações de conteúdo e contexto das fotografias.

Figura 8 – Visão geral do modelo



Fonte: Elaborado pelo autor

Na figura 8, tem-se uma visão geral do modelo proposto. Primeiramente, foi montada uma base de dados constituída de fotografias pessoais de diferentes usuários. Em seguida, foram gerados clusters para cada evento pessoal detectado nas fotografias. Logo após, houve a detecção dos eventos compartilhados, sendo realizada uma verificação de similaridade entre os eventos pessoais detectados. Por fim, houve os processos de anotação automática e semiautomática nas fotografias para cada evento compartilhado. Para a anotação semiautomática de pessoas, foram utilizados estimadores de contexto (coocorrência, recorrência espacial e recorrência temporal) e de conteúdo (reconhecimento e detecção de faces) e foi gerada uma lista de recomendações para cada fotografia. Já para a anotação automática de pessoas, foram realizados a detecção e reconhecimento das faces nas fotografias e, em seguida, foi feita a propagação das anotações de pessoas.

É importante ressaltar que este trabalho consiste na junção de dois outros trabalhos: Figueirêdo et al. (2012), que propôs a detecção de eventos compartilhados em coleções de fotografias e Figueirêdo et al. (2015), que propôs técnicas de anotação automática e semiautomática de pessoas em fotografias. Como pode ser visto na figura 8, além da junção destes dois trabalhos, foram efetuadas modificações na etapa de detecção e reconhecimento de faces tanto na anotação automática de pessoas, quanto na semiautomática.

4.1 Detecção de Eventos Compartilhados

Como mencionado inicialmente, um evento pode ser definido como um acontecimento que reúne pessoas com algum objetivo em comum durante um período de tempo e em um determinado lugar. Exemplos de eventos incluem festas de aniversário, shows e casamentos.

Para realizar a anotação de eventos nas fotografias, foi utilizado um algoritmo automático (porém configurável) de detecção de eventos nas fotografias. O algoritmo se utiliza de informações contextuais das fotografias, adquiridas por meio dos metadados - data e hora da captura e localização geográfica da fotografia.

Esta seção trata da detecção de eventos compartilhados, que decorrem da existência de vários dispositivos fotográficos (câmeras ou smartphones) de diversas pessoas em um mesmo evento. Geralmente, cada pessoa possuidora de um dispositivo possui sua organização peculiar de fotografias. Como consequência do fato de dispositivos distintos possuírem funções diferentes, é possível que fotografias capturadas num mesmo evento possuam tipos de informações nos metadados singulares. Por exemplo, as fotografias capturadas a partir da câmera de um smartphone, equipada com GPS, possuem a informação de localização geográfica nos seus metadados, enquanto que em outra câmera, equipada com detector de faces, as fotografias capturadas possuem

apenas as informações das faces detectadas.

Além disso, é possível que algumas fotografias tenham seus metadados perdidos ou alterados. Ao adicionar fotografias em redes sociais - como Facebook, por exemplo - as informações de localização geográfica são removidas. Assim, uma fotografia pode ter sido anotada como participante de um evento, porém, sem as informações temporais e/ou espaciais. Isto torna evidente a importância do compartilhamento de anotações entre fotos de um mesmo evento compartilhado.

Com a detecção de eventos compartilhados, pretende-se que anotações de pessoas ocorram de forma automática ou semiautomática. Por exemplo, João foi anotado em um evento como presente e em outro evento essa anotação não constava. Se os dois eventos são tidos como um evento em comum, a propagação da anotação do primeiro evento é realizada para o segundo.

4.1.1 Detecção de Eventos Pessoais

A organização de fotografias em eventos pode envolver dois problemas: a identificação do evento no 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. No entanto, este trabalho trata da anotação de eventos e não do agrupamento de fotografias.

Para realizar a detecção de eventos pessoais, é necessário que as fotografias estejam anotadas com as informações da localização geográfica, e data e hora de captura. Neste momento, é considerado que não existem eventos paralelos na coleção de fotografias de um mesmo usuário.

Primeiramente, o modelo reúne fotografias de diferentes usuários. Em seguida, a detecção de eventos pessoais é realizada. O modelo proposto considera as informações espacial e temporal das fotografias para criar clusters usando o algoritmo DBSCAN (ESTER et al., 1996). O DBSCAN verifica 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 todos os atributos de uma única vez ou para cada atributo separadamente, criando-se subgrupos. Levando em consideração o contexto de fotografias e as dimensões espacial e temporal, o agrupamento pode ser feito das seguintes maneiras:

- 1) Utilizando as duas dimensões;
- 2) Utilizando a dimensão temporal inicialmente e, para cada grupo da dimensão

temporal, um agrupamento espacial diferente;

- 3) Utilizando a dimensão espacial inicialmente e, para cada grupo da dimensão espacial, um agrupamento temporal diferente.

Como a informação temporal sempre está presente nas fotografias digitais, primeiramente é feito o agrupamento da dimensão temporal e, depois, da dimensão espacial separadamente. Assim, se uma fotografia não tem a localização geográfica anotada, ela vai ser associada somente ao cluster temporal criado pelo algoritmo.

O algoritmo DBSCAN requer dois parâmetros de entrada: epsilon (distância mínima entre os membros do mesmo cluster) e o número mínimo de objetos em cada cluster. Neste trabalho, o número mínimo de fotos por cluster pode ser 1, considerando que uma foto isolada pode ser considerado como um cluster. Lacerda et al. (2008a) realizaram experimentos para verificar o melhor epsilon no tocante à anotações de eventos em fotografias, tanto para a clusterização temporal, quanto para a espacial. Assim, neste trabalho utilizou-se os valores sugeridos pelos autores, que foram: 24 horas para o cluster temporal e 1 km para o cluster espacial.

A clusterização temporal cria conjuntos de fotos segmentados pelo tempo usando um tempo máximo como parâmetro. Deste modo, os clusters têm fotos com distâncias temporais menores ou iguais ao tempo máximo, e duas fotos pertencem ao mesmo conjunto, se a distância temporal entre elas é menor ou igual ao tempo máximo. A clusterização espacial é feita de forma semelhante à temporal porém, a distância entre as fotos é verificada pela distância geodésica a partir das coordenadas geográficas das fotografias.

4.1.2 Detecção de Eventos Compartilhados

Com relação à anotação de eventos, verificou-se que as redes sociais e as informações de contexto são as mais utilizadas para o agrupamento das fotografias em eventos. Porém, o fato de várias pessoas terem capturado fotografias de diferentes dispositivos fotográficos e armazenado em locais distintos não tem sido considerado, apesar de essa informação poder ser utilizada para a anotação de fotografias de diferentes usuários.

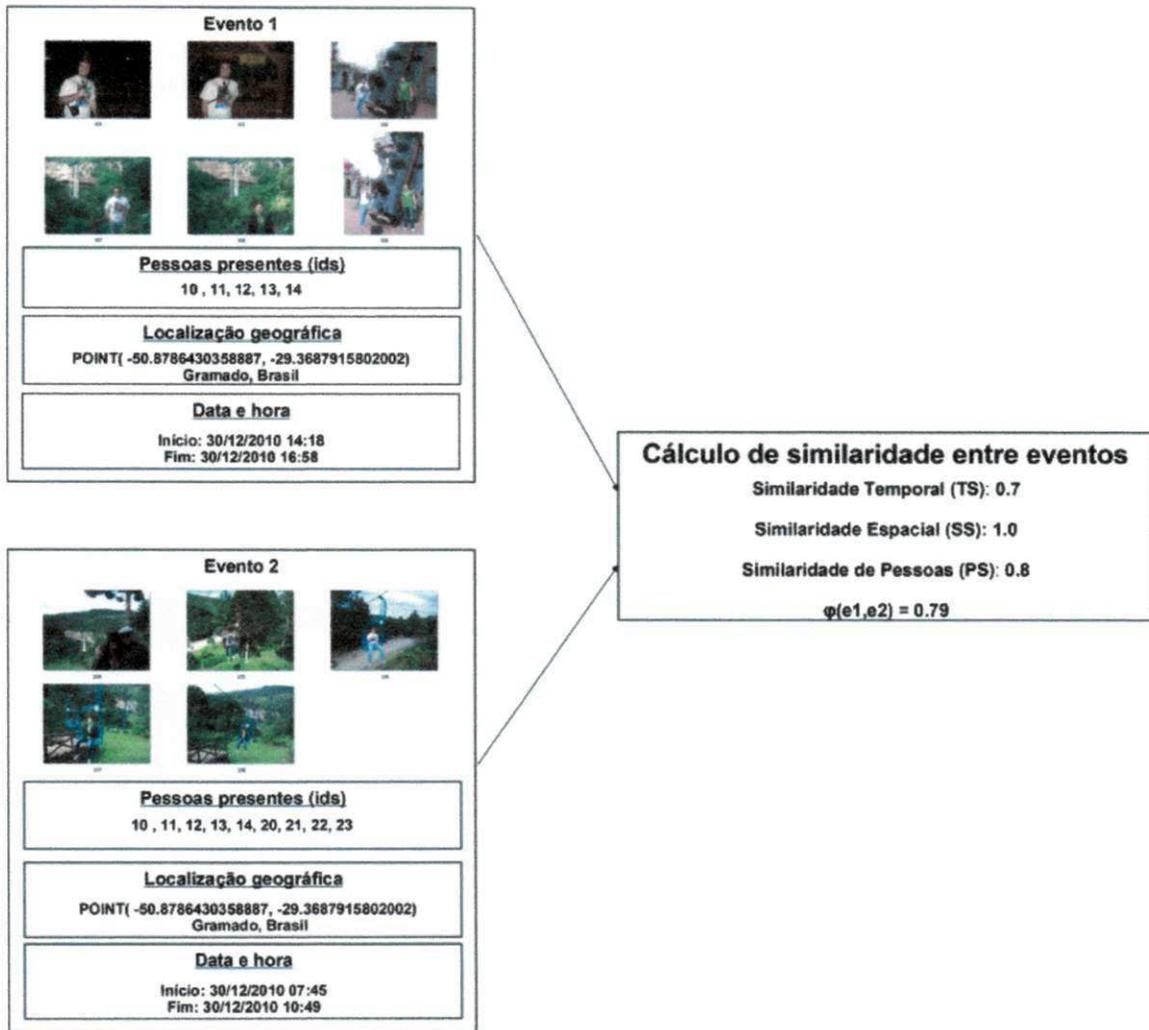
O algoritmo de detecção de eventos pessoais obteve bons resultados para coleções de fotografias pessoais porém, mostrou-se ineficaz para a detecção de eventos compartilhados. O principal motivo disto é a existência de eventos que ocorrem ao mesmo tempo nas coleções de fotografias dos usuários. Para detectar os eventos compartilhados, utilizou-se o algoritmo proposto por Figueirêdo et al. (2012).

O algoritmo utilizado nesta pesquisa se baseia no cálculo de similaridade entre pares de eventos, considerando a distância geográfica, a distância temporal e as pessoas presentes nos eventos. Para cada par de eventos, a função de similaridade temporal calcula a interseção entre o tempo inicial e final dos dois eventos; a função de similaridade espacial, por sua vez, calcula a interseção entre a localização geográfica das fotografias dos eventos e a função de similaridade de pessoas calcula a interseção de pessoas presentes nos eventos. Cada função de similaridade retorna um valor real entre 0 e 1, indicando o quão similares são os dois eventos passados como parâmetro.

A figura 9 mostra um exemplo do cálculo de similaridade entre dois eventos: os eventos 1 e 2. O evento 1 aconteceu no dia 30/12/2010, iniciando às 14:18 e terminando às 16:58. Já o 2, que ocorreu no mesmo dia, teve início às 07:45 e término às 10:49. Ambos os eventos se deram na mesma localização geográfica (mesmo par de coordenadas): Gramado, Brasil. Todas as pessoas presentes no evento 1 também estiveram presentes no 2; porém, este último conteve quatro pessoas a mais do que o outro. Ao observar as fotos tomadas como amostra (presentes na figura 9), a localização geográfica, as pessoas presentes e as datas de ambos os eventos, pode-se perceber que se trata de um evento compartilhado.

Para verificar se os eventos 1 e 2 são o mesmo evento compartilhado, foi feito o cálculo da similaridade entre os dois eventos. Primeiramente, faz-se necessário atribuir valores aos pesos para as similaridades espacial, temporal e de pessoas. Como já mencionado anteriormente, a informação temporal sempre está presente em uma fotografia; enquanto as pessoas e a localização geográfica podem não estar. Por esta razão, foi atribuído um valor maior para o peso da função de similaridade temporal - 0.6. Os pesos das funções de similaridade geográfica e de pessoas terão o mesmo valor - 0.2. Como se pode ver na figura 9, as funções de similaridade espacial, temporal e de pessoas tiveram como resultados os valores 1.0, 0.7 e 0.8, respectivamente. É importante ressaltar que o intervalo de tempo adicional utilizado na função de similaridade espacial foi de 4 horas (FIGUEIRÊDO et al., 2012). Por fim, com os valores das funções de similaridade temporal, espacial e de pessoas, e com os respectivos pesos destas funções, pode-se obter o valor final da função que calcula a similaridade entre dois eventos ($\varphi(e1, e2)$) (equação 2.2). Pode-se observar, de acordo com a figura 9, que a função de similaridade teve como resultado o valor 0.79. Com este valor, utilizando-se como limiar 0.7, pode-se dizer que os eventos 1 e 2 fazem parte de um mesmo evento compartilhado.

Figura 9 – Exemplo de cálculo de similaridade entre dois eventos



Fonte: Elaborado pelo autor

4.2 Anotação Automática de Pessoas em Fotografias

A anotação de pessoas nas fotografias auxilia o usuário na recuperação e organização delas. Uma das técnicas mais utilizadas para anotação automática de pessoas é baseada na análise do conteúdo da imagem (HUANG; JIAU; JIAN, 2016; SHIMIZU; NITTA; BABAGUCHI, 2011; STONE; ZICKLER; DARRELL, 2008). Por meio desta análise, algumas características importantes da imagem podem ser identificadas para auxiliar na anotação de pessoas.

Utilizando-se da análise de conteúdo, inicialmente deve ser feita a detecção das faces presentes nas fotografias. Esta tarefa consiste em identificar os componentes da face humana na fotografia - como olhos, boca e nariz.

Em seguida, as faces detectadas anteriormente devem ser identificadas. Para tal, o reconhecimento de faces é realizado, permitindo que as pessoas presentes na fotografia sejam identificadas. O reconhecedor utiliza algumas características detectadas nas faces, como distância entre os olhos, por exemplo. Essas características são comparadas com as das pessoas contidas nas fotografias da base de treinamento.

4.2.1 Detecção de faces

Para realizar a tarefa de detecção de faces, foi utilizada a solução proposta por Markus et al. (2013). Após as regiões que representam faces, segundo o classificador, serem detectadas nas fotografias, estas são inseridas no banco de dados. Após isso, é feito o reconhecimento de todas as faces detectadas.

4.2.2 Reconhecimento de faces

O passo seguinte após a detecção de faces é o reconhecimento das regiões armazenadas no banco de dados. Neste trabalho, para realizar o reconhecimento de faces foi utilizada uma versão modificada do algoritmo de EigenFaces (TURK; PENTLAND, 1991), o qual utiliza a distância euclidiana entre elementos presentes na face humana para atribuir uma face à uma pessoa específica. No algoritmo original, para cada pessoa existente na base de dados, fazia-se necessário que esta tivesse sido anotada ao menos uma vez; assim, o algoritmo conseguia gerar um modelo com base nas faces anotadas de cada pessoa. No código 4.1, é possível visualizar um exemplo de trecho de código do algoritmo EigenFaces, no qual se pode ver que são obtidas todas as faces conhecidas dos contatos de um determinado usuário. Este usuário é passado como parâmetro no construtor da classe. Deste modo, o algoritmo trabalha apenas com todas as faces conhecidas dos contatos de um determinado usuário.

Código 4.1 – Exemplo de trecho do algoritmo de EigenFaces

```
1 int numTrainingImages = knownFaces.size();
```

No entanto, de acordo com a abordagem proposta, dentro de um evento compartilhado pode haver pessoas que ainda não foram anotadas por um determinado usuário; porém, podendo ter sido anotada por outro usuário distinto. Desse modo, deveria ser possível haver a propagação de anotações dentro de um mesmo evento compartilhado. Para tanto, foi necessário realizar algumas modificações no algoritmo de EigenFaces.

Primeiramente, anotações para cada pessoa existente na base de dados passaram a não ser mais obrigatórias. Destarte, o reconhecedor de faces passou a aceitar faces não identificadas previamente. Junto com a face a ser reconhecida, também é recebido um conjunto de faces (do conjunto de treinamento) com as quais é feita uma comparação, de modo a identificar as mais semelhantes. Por fim, as faces mais

semelhantes à face a ser reconhecida são retornadas, e o algoritmo seleciona a mais semelhante para realizar a anotação automática.

No código 4.2 (das linhas 23 a 58), pode-se observar as alterações que foram feitas no algoritmo, a fim de que fosse possível a utilização do conceito de evento compartilhado. Na versão original do código, o método recebia apenas uma região (face) como parâmetro, e dentre os contatos de um determinado usuário (o qual é passado como parâmetro no construtor da classe) era buscado o usuário cuja face fosse mais semelhante à passada como parâmetro. Na versão modificada, uma lista de fotografias é passada como parâmetro, a fim de se encontrar a face mais semelhante à passada como parâmetro dentre as que foram detectadas nas fotografias.

Código 4.2 – Alteração no código do algoritmo de EigenFaces

```

1 region , List<Photo> names) throws Exception {
2     BufferedImage unknownFaceImage = RegionUtils.toThumbnail(
3         region ,
4         Constants.FACE_THUMBNAIL_SIZE);
5     byte[] unknownFace =
6     Utilities.bufferedImageToIntensityArray(unknownFaceImage);
7     double[] unknownFaceWeight =
8     eigenfaceRecognition.getWeightForImage(unknownFace);
9     double[] unknownMirroredFaceWeight = null;
10    byte[] mirroredFace = new byte[unknownFace.length];
11    for (int i = 0; i < Constants.FACE_THUMBNAIL_SIZE.height; i
12        ++){
13        for (int j = 0; j < Constants.FACE_THUMBNAIL_SIZE.
14            width; j++){
15            int elem = i * Constants.FACE_THUMBNAIL_SIZE.
16                width + j;
17            mirroredFace[elem] = unknownFace[(i + 1) *
18                Constants.FACE_THUMBNAIL_SIZE.width - j - 1];
19        }
20    }
21    unknownMirroredFaceWeight =
22    eigenfaceRecognition.getWeightForImage(mirroredFace);
23    HashMap<Integer, Double> result = new HashMap<Integer, Double>
24    (>names.size());
25    ArrayList<SortableContainer<Region>> bestHits = new
26    ArrayList<SortableContainer<Region>>();
27    for (Photo p : names) {
28        List<Region> regioesDaFoto =
29        PhotoGeoController.getInstance().getRegionsForImage(p
30            .getId());
31        for (int i = 0; i < regioesDaFoto.size(); i++) {
32            Region image = null;
33            List<Region> regionsForName =
34            PhotoGeoController.getInstance()
35            .getRegionsForFace(regioesDaFoto.get(i).
36                getContactId());
37            if (regionsForName != null) {
38                double minDist = Double.MAX_VALUE;
39                for (int j = 0; j < regionsForName.
40                    size(); j++) {

```

```

32         if (regionsForName.get(j) !=
33             null &&
34             regionsForName.get(j) !=
35             region) {
36             byte[] knownFace =
37                 Utilities.bufferedImageToIntensityArray(RegionUtils.toThumbnail
38                 (regionsForName.get(j), Constants.FACE_THUMBNAIL_SIZE));
39             double[]
40                 knownFaceWeight =
41                 eigenfaceRecognition.getWeightForImage(knownFace);
42             double distance =
43                 eigenfaceRecognition
44                 .getDistanceBetweenWeights(knownFaceWeight, unknownFaceWeight);
45             if (
46                 unknownMirroredFaceWeight
47                 != null) {
48                 distance =
49                     Math.min(
50                         distance,
51                         eigenfaceRecognition.getDistanceBetweenWeights(knownFaceWeight,
52                         unknownMirroredFaceWeight));
53                 if (distance <
54                     minDist) {
55                     minDist =
56                         distance;
57                     image =
58                         regionsForName
59                         .get(j);
60                 }
61             }
62         }
63     }
64     result.put(regionsForName.get(i).
65         getContactId(), minDist);
66     if (image != null && minDist != 0) {
67         bestHits.add(new
68             SortableContainer<Region>(
69                 image, 56: minDist));
70     }
71 }
72 }
73 }
74 }
75 }
76 }
77 }
78 }
79 }
80 }
81 }
82 }
83 }
84 }
85 }
86 }
87 }
88 }
89 }
90 }
91 }
92 }
93 }
94 }
95 }
96 }
97 }
98 }
99 }
100 }

```

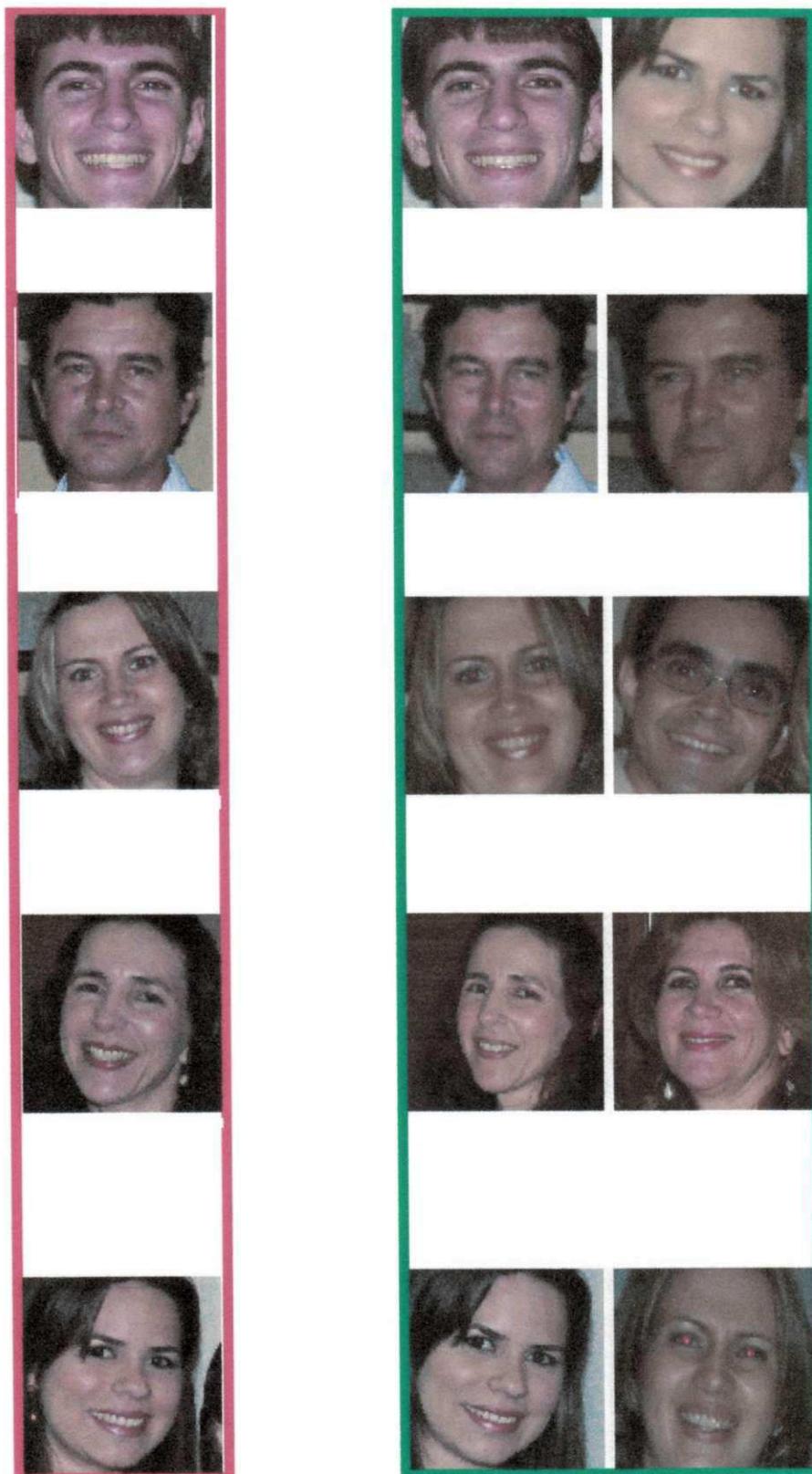
Como pode ser visto no código 4.2, para cada face detectada no conjunto de fotografias passado como parâmetro, são feitas comparações para encontrar a face cuja distância da face passada como parâmetro seja a menor possível. É computada uma pontuação para cada face e, ao final, uma lista com todas as pontuações é retornada para posteriormente haver a seleção da face mais semelhante à que foi passada como parâmetro.

Esta customização no EigenFaces foi necessária para permitir que faces não

anotadas na coleção de fotografias de um usuário pudessem ser anotadas com base nas faces mais similares presentes nas coleções de fotografias de outros usuários que também estiveram presente no mesmo evento compartilhado.

Na figura 10, pode-se ver exemplos de faces detectadas e as mais semelhantes retornadas pelo algoritmo de EigenFaces modificado.

Figura 10 – À esquerda (quadro vermelho) estão as faces detectadas pelo algoritmo; à direita (quadro verde) estão as faces mais semelhantes às respectivas faces detectadas.



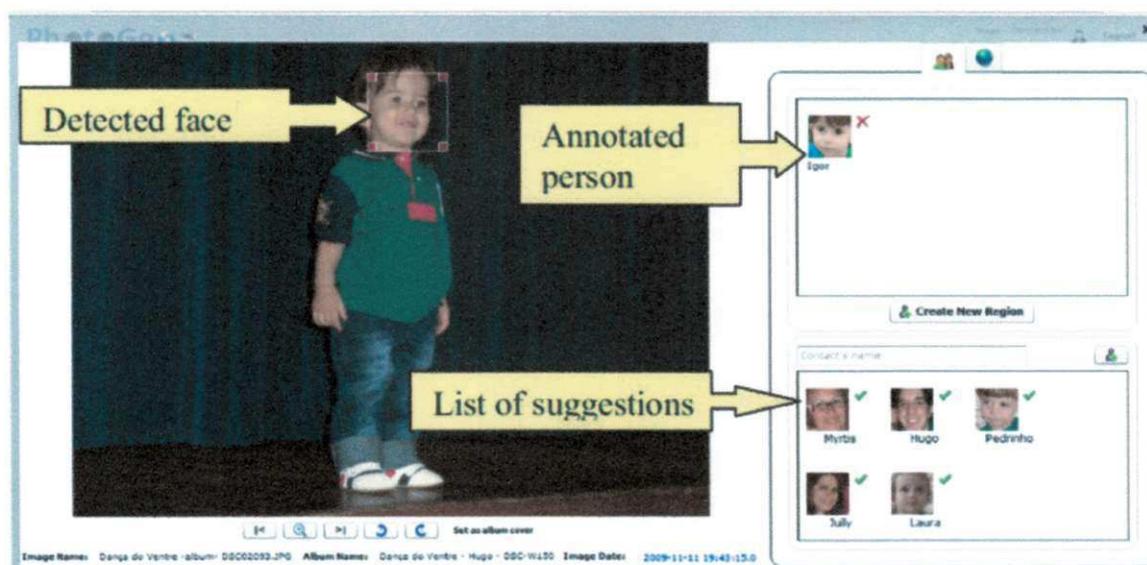
Fonte: Elaborado pelo autor

Como pode ser visto na figura 10, à esquerda - destacado em vermelho - têm-se as faces detectadas pelo algoritmo de EigenFaces modificado. À direita - destacado em verde - são exibidas as faces detectadas mais semelhantes dentre as do conjunto de treinamento, utilizando o conceito de evento compartilhado. Como mencionado anteriormente, sob essa abordagem não é necessário que as faces sejam identificadas previamente para a execução do algoritmo de reconhecimento de faces.

4.3 Anotação Semiautomática de Pessoas em Fotografias

Para realizar a anotação semiautomática de pessoas nas fotografias, foi utilizada uma técnica baseada em uma soma ponderada de estimadores, proposta por Figueirêdo et al. (2015). Estes estimadores fazem uso do contexto (informações temporais, por exemplo) e do conteúdo (faces detectadas, por exemplo) das fotografias e avaliam a probabilidade de um contato $c \in C$ (onde C é a lista completa de contatos do usuário) estar presente na fotografia p . Por fim, os estimadores geram uma lista de recomendações com as pessoas mais propensas a estar na fotografia p (figura 11).

Figura 11 – Solução proposta por Figueirêdo et al., 2015



Fonte: Figueirêdo et al., 2015

Os estimadores de contexto utilizados nesta pesquisa foram os seguintes:

- **Recorrência Espacial:** indica a porcentagem de vezes que um contato foi anotado em uma fotografia dentro de um raio de d metros da localização geográfica da fotografia a ser anotada;
- **Recorrência Temporal:** indica a porcentagem de vezes que um contato foi anotado num intervalo t do momento em que a fotografia a ser anotada foi capturada;

- **Coocorrência:** indica a porcentagem de vezes que um contato está presente em uma fotografia simultaneamente com outro contato que já foi anotado naquela fotografia.

Além dos estimadores de contexto, também foi utilizado um estimador de conteúdo: reconhecimento de faces. Este estimador indica o contato com maior pontuação entre os valores retornados pelo algoritmo de reconhecimento de faces aplicado às faces detectadas na fotografia. Os algoritmos de reconhecimento e detecção de faces foram os mesmos utilizados na etapa de anotação automática de pessoas.

4.4 Considerações Finais

Neste capítulo, foi proposto um algoritmo de anotação automática e semiautomática de pessoas em coleções de fotografias utilizando o conceito de evento compartilhado.

O referido algoritmo se divide em duas etapas após a detecção dos eventos compartilhados: anotação automática e semiautomática de pessoas. Na anotação semiautomática, o usuário é auxiliado por uma lista de sugestões com as pessoas mais prováveis de estarem presentes nas fotografias. Essa lista é gerada a partir da ordenação decrescente do valor obtido a partir da soma ponderada de estimadores, que utilizam informações contextuais e de conteúdo para calcular suas estimativas.

Na anotação automática de pessoas, são realizadas a detecção e o reconhecimento das faces das pessoas presentes nas fotografias. O processo de reconhecimento de faces foi customizado para permitir que faces previamente não anotadas (e não conhecidas) pudessem ser anotadas com base nas faces mais similares encontradas em coleções de fotografias de outros usuários que fizeram parte do mesmo evento compartilhado.

5 Avaliação e Resultados

Neste capítulo são apresentadas as bases de dados utilizadas neste trabalho, a metodologia empregada para avaliar o modelo proposto e os resultados obtidos. Inicialmente, é apresentada uma breve discussão concernente as bases de fotografias públicas pesquisadas e a base que foi utilizada neste trabalho, além de justificar o uso dela (seção 5.1). Em seguida, na seção 5.2, a metodologia utilizada para avaliar o modelo proposto para anotação automática e semiautomática de pessoas utilizando evento compartilhado é descrita. Por fim, na seção 5.3, os resultados obtidos da execução dos experimentos são exibidos e discutidos; e na seção 5.4 são apresentadas as considerações finais do capítulo.

5.1 Bases de dados utilizadas

Não foi encontrada uma base de fotografias pública que contenha a informação das pessoas presentes e da localização geográfica em seus metadados. As bases de fotografias públicas contém apenas a anotação das pessoas presentes (GALLAGHER; CHEN, 2008) ou apenas a localização geográfica das fotografias (IONESCU et al., 2016).

Varshney (2008) realizou um survey sobre a anotação da identidade das pessoas em coleções de fotografias. O autor argumenta que, apesar da importância de se inserir tags nas fotografias, as quais auxiliam os usuários na organização e gerenciamento de coleções de fotografias, existem questões de privacidade envolvendo este tema. Ele pesquisou trabalhos que relataram que cerca de 72% das fotos do Flickr tinham as tags marcadas como privadas, já que as pessoas podem marcar palavras relacionadas à sua localização nas fotografias.

Gallagher e Chen (2008) disponibilizaram uma base pública com 589 fotografias, anotadas apenas com as informações das pessoas presentes - ao todo, a coleção apresenta 32 pessoas (figura 12). A maioria das fotos tinha as faces bem iluminadas e todas as fotos foram capturadas utilizando apenas uma câmera. Inicialmente, essa base de fotografias foi utilizada com o propósito de verificar se a utilização das roupas usadas iria melhorar a precisão do reconhecimento de pessoas nas fotografias.

Figura 12 – Exemplos de imagens da base de dados de Gallagher e Chen (2008)



Fonte: Adaptado de Gallagher e Chen (2008)

No ano seguinte, os mesmos autores disponibilizaram uma base de fotografias obtidas do Flickr, composta por 5.080 imagens de grupos de pessoas com 28.231 pessoas anotadas, sendo a maior base deste tipo disponibilizada publicamente (GALLAGHER; CHEN, 2009). Os autores utilizaram esta base inicialmente com o intuito de estudar sobre a interseção entre o comportamento social e a visão computacional. Por exemplo, em fotos com grupos de pessoas, estas geralmente escolhem onde ficar com base em questões sociais (por exemplo, próximo a alguém significativo) ou em questões físicas (por exemplo, homens mais altos ficam na fila de trás). A figura 13 mostra alguns exemplos de fotografias desta base de dados.

Figura 13 – Exemplos de imagens da base de dados de Gallagher e Chen (2009)



Fonte: Adaptado de Gallagher e Chen (2009)

O MediaEval¹ é uma iniciativa de benchmarking dedicada à avaliação de novos algoritmos de acesso e recuperação multimídia. Ele enfatiza o 'multi' em multimídia e se concentra nos aspectos humanos e sociais de tarefas multimídia. A cada ano, são disponibilizadas bases de dados para realização de tarefas propostas nesse evento - a exemplo da "Placing Task", que é uma tarefa recorrente no MediaEval, cujo objetivo é estimar a localização de fotos e vídeos disponibilizados em redes sociais apenas analisando o conteúdo e os metadados destes arquivos.

As bases de dados de fotografias disponibilizadas pelo MediaEval não contêm anotações de pessoas, embora contenham pessoas nas fotografias. Este fato ocorre

¹ <http://www.multimediaeval.org/>

porque o propósito das tarefas do MediaEval não tem a ver com anotações de pessoas. Como exemplos de bases de dados do MediaEval tem-se a disponibilizada por Ionescu et al. (2016), a qual foi disponibilizada para auxiliar na resolução da tarefa que consistia em recuperação de imagens sociais. A base de fotografias continha cerca de 87.000 fotos obtidas do Flickr e estava anotada com a localização geográfica e a informação temporal, além de outros metadados, como número de visualizações da foto, por exemplo. Conci et al. (2015) disponibilizaram uma base de fotografias consistindo em 2124 fotografias obtidas dos Jogos Olímpicos de Verão, que ocorreram em Londres de 2012; e 1351 fotografias obtidas dos Jogos Olímpicos de Inverno, que ocorreram em Vancouver. Todas as fotografias estavam anotadas com a informação temporal e algumas, com a localização geográfica. Na figura 14, pode-se ver alguns exemplos de fotografias desta base de dados.

Figura 14 – Exemplos de imagens da base de dados de Conci et al. (2015)



Fonte: Adaptado de Conci et al. (2015)

Além dos trabalhos citados acima, a maior parte dos trabalhos listados nesta

pesquisa também utilizaram coleções de fotografias privadas para realizar anotações de pessoas. Naaman et al. (2005) utilizou coleções de fotos pessoais em sua pesquisa, nas quais os usuários informavam a localização geográfica das fotografias. Jamil e Sa'dan (2014) utilizaram uma coleção pessoal de fotografias, composta por 174 fotografias que continham 138 faces de apenas três mulheres. Choi et al. (2011) optaram por construir uma base com fotos pessoais obtidas da rede social Cyworld (uma rede social coreana). Eles conseguiram coletar 547.991 imagens e contaram com a ajuda de voluntários para anotar as faces presentes.

Kim, El-Saddik e Jung (2012) utilizaram uma base de dados de fotos pessoais com 564 fotografias, construída por 26 participantes e anotada com as informações das pessoas presentes. O'Hare e Smeaton (2009) também utilizaram uma base de fotografias pessoais. Cerca de 24.000 fotografias foram enviadas por 29 usuários. No entanto, apenas as fotografias enviadas por nove destes usuários (os autores não citam o número exato de fotografias) tinham anotações de pessoas. Cooray e O'Connor (COORAY; O'CONNOR, 2009; COORAY, 2008; SADLIER et al., 2008) utilizaram uma base construída anteriormente pelo mesmo grupo de pesquisa (COORAY et al., 2006), consistindo em 14.000 fotografias pessoais enviadas por 22 usuários. Os usuários realizavam a anotação de pessoas utilizando o sistema construído no trabalho (MediaAssist).

Apenas dois dos trabalhos relacionados utilizaram bases públicas para validarem seus resultados. Escalante, Montes-y-Gómez e Sucar (2012) utilizaram uma base pública de imagens (chamada Corel) que continha imagens diversas - incluindo paisagens e animais. Esta base não pôde ser utilizada neste trabalho, pois ela não contém pessoas em suas fotografias. Brenner, Mirza e Izquierdo (2014) utilizaram a base disponibilizada por Gallagher e Chen (2008) em sua pesquisa, a qual é composta por 589 fotos pessoais com as anotações respectivas das pessoas presentes.

Dentre os trabalhos pesquisados, não foi encontrada nenhuma base de dados pública que satisfizesse as condições necessárias para que os experimentos realizados nesta pesquisa pudessem ser executados – isto é, as informações de tempo, localização geográfica e pessoas anotadas. Kim, El-Saddik e Jung (2012) pontuaram que devido a algumas preocupações com a privacidade, era difícil conseguir muitos voluntários. Assim sendo, os autores não conseguiram coletar fotos e anotações em grande escala. Essa é uma das dificuldades para se encontrar uma base de dados pública que contenha anotações de pessoas e localização geográfica. As bases disponibilizadas por Gallagher e Chen (GALLAGHER; CHEN, 2008; GALLAGHER; CHEN, 2009) contém as pessoas anotadas porém, não dispõem das informações espaciais, as quais são imprescindíveis para a execução dos algoritmos utilizados nesta pesquisa, os quais separam as fotografias em eventos pessoais e verificam a similaridade entre estes

eventos. Por estas razões, optou-se por utilizar uma coleção de fotografias pessoais com todas as informações necessárias (data e hora de captura, localização geográfica e pessoas anotadas).

Assim sendo, nesta pesquisa foi utilizada uma coleção com 8.306 fotografias com 865 pessoas diferentes tanto para os experimentos com anotação semiautomática quanto para os com anotação automática de pessoas. As fotografias estavam anotadas com as informações de data e hora da captura, 89% delas anotadas com a localização geográfica e 68% anotadas com a informação das pessoas presentes - os próprios usuários que submeteram as fotografias as anotaram com as pessoas presentes. Tendo em vista que esta base de fotografias foi composta por fotos enviadas por diversas pessoas, não foi possível conseguir o consentimento de todos os envolvidos para disponibilizar esta base de fotografias publicamente.

A base de fotografias disponibilizada por Gallagher e Chen (2008) foi utilizada nos experimentos com anotação automática de pessoas, tão somente com o fim de verificar se o desempenho do algoritmo de detecção e reconhecimento de faces era satisfatório para uma base de dados diferente. No entanto, não foi possível utilizar esta base neste trabalho para verificar se o conceito de evento compartilhado traz algum benefício na anotação automática e semiautomática de pessoas, pois ela não era dotada das informações espaço-temporais (imprescindíveis para a separação das fotografias em eventos). Por isso, optou-se por utilizar uma base privada de fotos, contendo todas as informações necessárias à execução dos experimentos.

5.2 Metodologia de Avaliação

O algoritmo de detecção de eventos pessoais foi executado para cada coleção de fotografias, dividindo-as em 298 eventos pessoais. Após isso, o algoritmo de detecção de eventos compartilhados foi executado para cada par de eventos pessoais. Um par de eventos pessoais e_1 e e_2 é considerado como um evento compartilhado se o resultado da função $\varphi(e_1, e_2)$ (descrita na equação 2.2) for maior ou igual a 0,8 (FIGUEIRÊDO et al., 2012). A coleção de fotografias utilizada aqui foi derivada de Figueirêdo et al. (2015). Eles fizeram experimentos e constataram que os melhores valores para os pesos dos estimadores de similaridade espacial, temporal e de pessoas (mostrados na equação 2.2) foram de 0,2, 0,6 e 0,2, respectivamente. Destarte, considerou-se os mesmos valores para esses pesos neste trabalho. Dentre todos os eventos pessoais, a execução da função φ resultou em 53 eventos compartilhados. Dado que um evento compartilhado é composto de, pelo menos dois eventos pessoais, dos 298 eventos pessoais iniciais, no mínimo 106 deles puderam ser agrupados em algum evento compartilhado.

A métrica utilizada para realizar a comparação dos resultados foi a taxa de

acerto, que é a razão entre o número de acertos e o número de pessoas presentes na fotografia, sendo utilizada para mensurar a eficiência de técnicas que sugerem listas de pessoas. No caso da anotação semiautomática, foram realizados experimentos com vários tamanhos de listas de pessoas - entre 1 e 10. Esta métrica foi utilizada neste trabalho, pois tem sido utilizada em vários trabalhos relacionados (FIGUEIRÉDO et al., 2015; CHOI et al., 2011; O'HARE; SMEATON, 2009). A taxa de acerto para cada fotografia foi calculada como pode ser visto na equação 5.1:

$$H_p = \frac{n_h}{n_f} \quad (5.1)$$

Onde n_h é o número de acertos do algoritmo e n_f é o número de faces presentes na fotografia. Por sua vez, a taxa de acerto por evento foi calculada utilizando a soma das taxas de acerto de todas as fotografias do evento (equação 5.2):

$$H_e = \sum_{h \in H_p} h \quad (5.2)$$

Onde $\frac{\sum_{h \in H_p} h}{n_p}$ é o somatório das taxas de acerto de todas as fotografias do evento e n_p é o número total de fotografias.

É importante ressaltar que a taxa de acerto foi calculada utilizando o teste *Gold Standard* ou padrão de ouro (LEONG; MIHALCEA; HASSAN, 2010), que consiste na comparação dos resultados obtidos no experimento com os dados reais - no caso deste estudo, anotações feitas pelos usuários manualmente.

5.2.1 Anotação automática de pessoas em fotografias

Foram conduzidos dois experimentos diferentes para a anotação automática de pessoas: um utilizando uma base pública de fotografias (disponibilizada por Gallagher e Chen (2008)), com o propósito de verificar a taxa de acerto do algoritmo de detecção e reconhecimento de faces empregado nesta pesquisa; e outro utilizando a base privada de fotos (descrita na seção 5.1), com o objetivo de verificar se o conceito de evento compartilhado traz algum incremento na taxa de acerto na anotação automática de pessoas em fotografias.

Para o experimento utilizando a base privada de fotografias, após a detecção dos eventos compartilhados, foi utilizada a técnica k-fold cross validation (k=10) para verificar o impacto desse tipo de evento na anotação de pessoas. Assim, para cada evento compartilhado todas as fotos foram divididas em dez subgrupos, e em cada iteração, um deles foi selecionado como um grupo de teste e os restantes compunham o conjunto de treinamento. Para cada face detectada em cada foto do grupo de treinamento, era

feita uma comparação com todas as faces detectadas nas fotos do grupo de teste para verificar qual face seria mais semelhante. Caso a face mais semelhante encontrada no grupo de teste fosse da mesma pessoa da face detectada do grupo de treinamento, um acerto era contabilizado.

O objetivo deste experimento foi verificar o impacto dos eventos compartilhados na anotação automática de pessoas. Primeiro, apenas algumas pessoas foram anotadas nas coleções de fotos. Em seguida, foi executada a anotação automática de pessoas para as fotografias de cada evento compartilhado, utilizando validação cruzada.

Para verificar o impacto de utilizar o conceito de evento compartilhado sobre a anotação automática de pessoas, foi realizada uma comparação entre os resultados da anotação automática utilizando evento compartilhado e os resultados sem utilizar evento compartilhado.

5.2.2 Anotação semiautomática de pessoas em fotografias

De acordo com Figueirêdo et al. (2015), a utilização dos estimadores de conteúdo e contexto das fotografias obtiveram melhores resultados do que a utilização destes tipos em separado. Por isso, neste trabalho foram usados os estimadores de coocorrência, recorrência espacial, recorrência temporal e reconhecimento de faces.

Após a detecção dos eventos compartilhados, foi utilizada a técnica k-fold cross validation ($k=10$) a fim de verificar o impacto deste tipo de evento na anotação semiautomática de pessoas. Assim, para cada evento compartilhado todas as fotografias foram divididas em dez subconjuntos, e em cada iteração, um deles era selecionado como conjunto de teste e os restantes compunham o conjunto de treinamento. Para cada foto do conjunto de treinamento, as faces de todas as pessoas presentes eram detectadas. Para cada face detectada, era gerada uma lista de recomendações com as pessoas detectadas no conjunto de teste mais propensas de serem a pessoa cuja face foi detectada. Esta lista de recomendações era gerada com base em todos os estimadores discutidos previamente (coocorrência, recorrência temporal e espacial, e reconhecimento de faces). Caso a pessoa da face detectada estivesse na lista de recomendações, um acerto era contabilizado.

Para verificar o impacto dos eventos compartilhados sobre a anotação semiautomática de pessoas, foi realizada uma comparação entre os resultados da anotação com e sem a utilização de eventos compartilhados.

5.3 Resultados

Esta seção apresenta os resultados para os algoritmos propostos para anotação automática e semiautomática de pessoas. Para os dois tipos de anotação, como discutido anteriormente, executou-se o algoritmo proposto para os eventos compartilhados, foram obtidos os resultados e foi feita uma comparação com a execução do algoritmo para os mesmos eventos separadamente, sem levar em consideração o conceito de evento compartilhado.

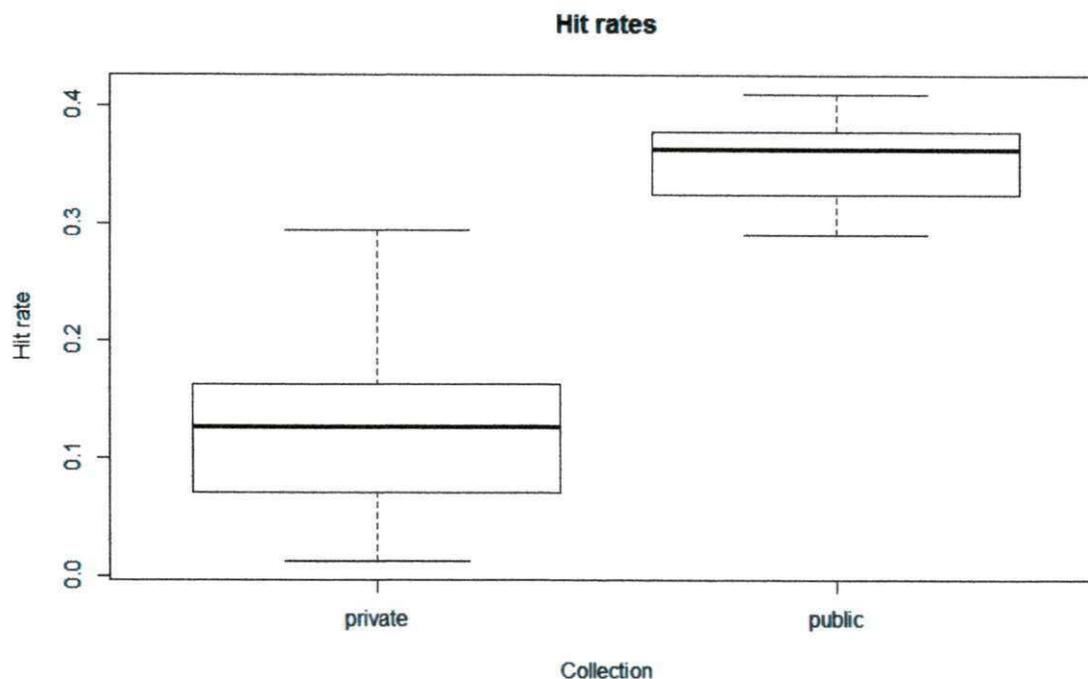
5.3.1 Anotação automática de pessoas em fotografias

Dois experimentos distintos foram realizados com anotação automática de pessoas. Como discutido na seção 5.1, não foi possível encontrar uma base pública com todas as informações necessárias para verificar o impacto da utilização do conceito de evento compartilhado. No entanto, obteve-se uma base de fotografias pública (GALLAGHER; CHEN, 2008), disposta tão somente das pessoas anotadas e ela foi utilizada para comparar os algoritmos utilizados para detecção e reconhecimento de faces na base privada de fotografias. Para a detecção de faces, foi utilizada a solução desenvolvida por Markus et al. (2014). O segundo experimento consistiu em verificar o impacto da utilização do conceito de evento compartilhado na anotação automática de pessoas utilizando uma base pessoal de fotografias (descrita na seção 5.1).

A figura 15 destaca a diferença entre os resultados da execução do algoritmo para a coleção usada (privada) e a coleção pública de fotos pública - disponibilizada por Gallagher e Chen (2008). É possível perceber que a coleção pública apresenta uma taxa de acerto significativamente maior que a utilizada neste trabalho, além de uma menor dispersão nos resultados. Isso acontece devido ao fato de que a grande maioria das fotos da base pública tinha as faces bem iluminadas, além do fato de que apenas uma câmera foi usada para capturar todas as fotografias dessa base, garantindo uma qualidade equivalente para todas as fotografias.

Outro ponto importante foi que na coleção pública de fotos, o conjunto de pessoas era pequeno (apenas 32 pessoas). Na maioria das fotos capturadas havia apenas seis pessoas - entre as 589 fotos - e houve uma pessoa que apareceu em 165 delas. Enquanto isso, na coleção privada de fotos, houve um evento compartilhado (composto por mais de um evento pessoal) com 165 fotos e 59 pessoas anotadas - a pessoa com mais incidência na coleção apareceu em apenas 30 fotos.

Figura 15 – Comparação da execução do algoritmo de detecção e reconhecimento de faces usando uma base pública e a base de fotos pessoais adotada



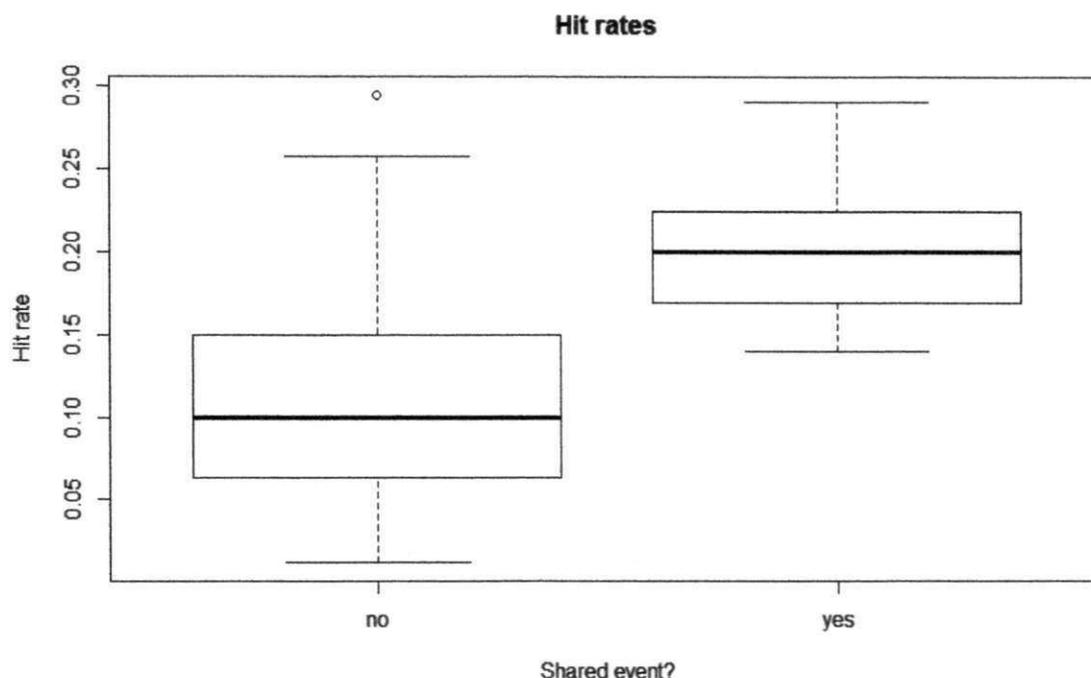
Fonte: Elaborado pelo autor

Conforme discutido anteriormente, não foi possível usar a coleção pública de fotos para realizar os experimentos utilizando o conceito de evento compartilhado, pois esta não dispunha da localização geográfica em seus metadados, os quais são essenciais para a detecção de eventos compartilhados do algoritmo utilizado nesta pesquisa. Além disso, as fotografias são distribuídas em eventos pessoais de acordo com a informação temporal e espacial e o modelo da câmera. Já que na coleção pública todas as fotos foram capturadas pela mesma câmera, esta distribuição não poderia ter sido executada. Assim, a taxa de acerto da base pública foi calculada considerando que todas as fotos pertenciam ao mesmo evento, o que não corresponde à realidade, além de ser uma comparação injusta, uma vez que os eventos compartilhados detectados na coleção privada têm no máximo 250 fotos.

A figura 16 apresenta os resultados da execução do algoritmo na coleção privada de fotos utilizada, com o fim de verificar o impacto do conceito de evento compartilhado sobre anotação automática de pessoas em fotografias (segundo experimento). Para cada evento compartilhado na coleção de fotos, o algoritmo foi executado considerando todas as fotos como sendo do mesmo evento; e então, o algoritmo foi executado para as fotos de cada evento pessoal separadamente. Como pode ser observado, o uso do conceito de evento compartilhado aumentou a taxa de acerto média para

a coleção de fotos usada. Além disso, pode-se observar que usando o conceito de evento compartilhado, o desvio padrão diminuiu significativamente. Já que não há sobreposição entre os boxplots, pode-se afirmar que os resultados utilizando o conceito de evento compartilhado foram superiores aos da não utilização deste conceito.

Figura 16 – Comparação de resultados de anotação automática de pessoas sobre o uso de eventos compartilhados



Fonte: Elaborado pelo autor

Além da análise gráfica, também foi realizada uma análise estatística para verificar se o conceito de evento compartilhado trouxe algum incremento na taxa de acerto da anotação automática de pessoas. Assim, as seguintes hipóteses foram testadas:

- H1-0: A utilização do conceito de evento compartilhado não melhora a taxa de acerto para a anotação automática de pessoas;
- H1-1: A utilização do conceito de evento compartilhado melhora a taxa de acerto para a anotação automática de pessoas;

Para verificar a normalidade dos dados, foram utilizados os testes de Anderson-Darling (ANDERSON, 2011). Assumindo uma confiança de 95%, verificou-se que os dados seguiam uma distribuição normal. Em seguida, foi utilizado o teste t-Student (CHUNG, 1946) para verificar as hipóteses supracitadas. A hipótese nula adotada é

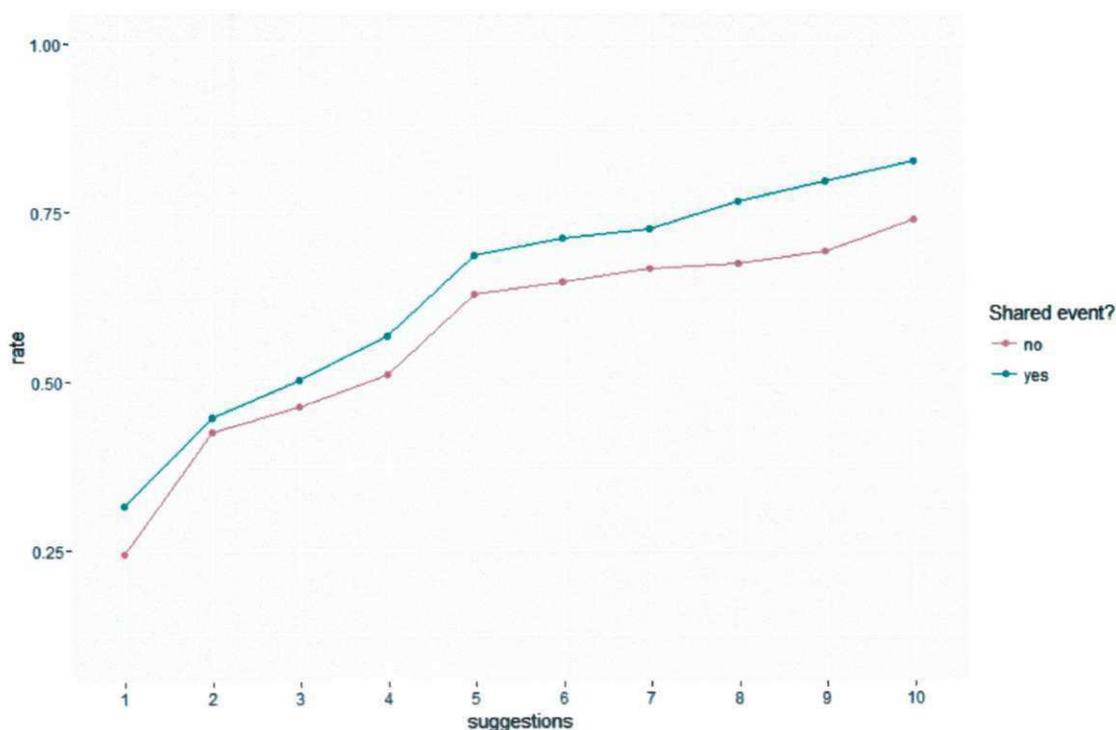
assumir que a utilização do conceito de evento compartilhado não melhora a taxa de acerto para a anotação automática de pessoas, ou seja, assumir que as médias dos resultados quanto a utilização de eventos compartilhados são iguais. Dado que o resultado do teste (cujo valor p foi menor que 0.05) dá suporte à rejeição da hipótese nula, pode-se afirmar estatisticamente que a utilização do conceito de evento compartilhado melhora a taxa de acerto para a anotação automática de pessoas.

5.3.2 Anotação semiautomática de pessoas em fotografias

Como discutido anteriormente, o algoritmo proposto foi executado para os eventos compartilhados, os resultados foram obtidos e comparados com a execução do algoritmo para os mesmos eventos separadamente, sem levar em consideração o conceito de evento compartilhado.

A figura 17 apresenta os resultados da execução do algoritmo para a coleção de fotografias utilizada, tendo em consideração o conceito de evento compartilhado. Para cada evento compartilhado na coleção de fotografias, o algoritmo foi executado considerando todas as fotos como participantes de um mesmo evento; e, em seguida, o algoritmo foi executado para as fotografias de cada evento separadamente. Como se pode notar, o uso do conceito de evento compartilhado aumentou a taxa de acerto em se tratando de anotação semiautomática de pessoas para a coleção de fotos utilizada.

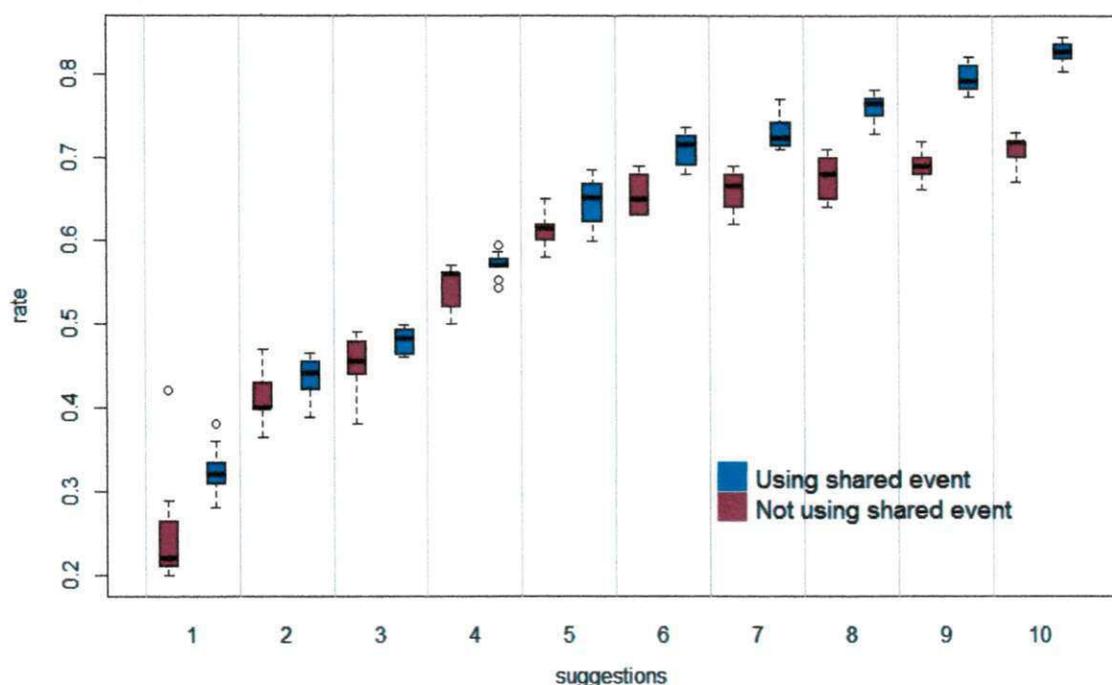
Figura 17 – Resultados da anotação semiautomática de pessoas considerando eventos compartilhados



Fonte: Elaborado pelo autor

A figura 18 apresenta os resultados da execução do algoritmo de anotação semiautomática de pessoas para a coleção de fotografias utilizada de forma mais detalhada, utilizando o detector de faces PICO. São apresentadas as distribuições dos resultados coletados utilizando validação cruzada para cada tamanho da lista de sugestões adotado. Como pode-se observar, para quase todos os casos - exceto quando a lista de sugestões teve tamanho 2 ou 3 -, não há sobreposição dos boxplots. Nestes casos, estatisticamente pode-se afirmar que os resultados utilizando o conceito de evento compartilhado foram superiores aos da não utilização deste conceito. Para os casos em que houve sobreposição nos boxplots (para os tamanhos 2 e 3), observa-se que não há sobreposição de medianas, o que possibilita afirmar que é muito provável que os resultados utilizando o conceito de evento compartilhado sejam superiores aos da não utilização deste conceito. No geral, pode-se considerar que os resultados que utilizaram o conceito de evento compartilhado foram superiores que os resultados que não levaram em consideração tal conceito. A Tabela 2 apresenta os resultados detalhados dos experimentos.

Figura 18 – Resultados detalhados da anotação semiautomática de pessoas considerando eventos compartilhados



Fonte: Elaborado pelo autor

Tabela 2 – Resultados detalhados da comparação entre anotação semiautomática de pessoas utilizando ou não o conceito de evento compartilhado

Número de recomendações	Taxa de acerto (utilizando evento compartilhado)	Taxa de acerto (sem utilizar evento compartilhado)
1	31%	24%
2	44%	42%
3	48%	47%
4	56%	57%
5	64%	62%
6	71%	64%
7	72%	66%
8	76%	67%
9	79%	69%
10	82%	73%

Fonte: Elaborado pelo autor

Além da análise gráfica, também foi realizada uma análise estatística para verificar se o conceito de evento compartilhado trouxe algum incremento na taxa de acerto da anotação semiautomática de pessoas. Assim, as seguintes hipóteses foram testadas:

- H2-0: A utilização do conceito de evento compartilhado não melhora a taxa de acerto para a anotação semiautomática de pessoas;

- H2-1: A utilização do conceito de evento compartilhado melhora a taxa de acerto para a anotação semiautomática de pessoas;

A normalidade dos dados foi verificada utilizando o teste de Anderson-Darling (ANDERSON, 2011). Com uma confiança de 95%, verificou-se que os dados não seguem uma distribuição normal. Foi então necessário utilizar um teste não-paramétrico para verificar as hipóteses acima. O teste Mann-Whitney (MANN; WHITNEY, 1947) foi utilizado para este fim. O resultado do teste (cujo valor p foi menor que 0.05) dá suporte à rejeição da hipótese nula; assim, é possível afirmar estatisticamente que a utilização do conceito de evento compartilhado melhora a taxa de acerto para a anotação semiautomática de pessoas.

5.4 Considerações Finais

Neste capítulo, foram discutidos em maiores detalhes os experimentos realizados sobre anotação automática e semiautomática de pessoas utilizando o conceito de evento compartilhado, incluindo a justificativa de escolha da base de fotografias utilizada, a metodologia empregada para avaliar o modelo proposto e os resultados obtidos ao longo da pesquisa.

De modo geral, os resultados foram satisfatórios, mostrando que a utilização do conceito de evento compartilhado trouxe um aumento na taxa de acerto na anotação automática e semiautomática de pessoas em fotografias.

No próximo capítulo, são apresentadas as considerações finais sobre o trabalho desenvolvido nesta pesquisa, suas contribuições e os trabalhos futuros.

6 Conclusões e Trabalhos Futuros

Neste capítulo são apresentadas as conclusões obtidas a partir dos resultados alcançados no desenvolvimento deste trabalho, sendo feito um paralelo com as hipóteses (seção 1.2) e os objetivos específicos (seção 1.1) adotados nesta dissertação.

A popularização de smartphones e dispositivos com câmeras fotográficas causou um grande aumento no número de fotografias gerado e armazenado pelos usuários. Este contingente de fotos torna difícil o processo de gerenciar e pesquisar nestes arquivos. Para viabilizar um gerenciamento mais eficaz, é imprescindível o uso de anotações nas fotografias. Neste trabalho, foi proposta uma solução para anotar pessoas e eventos; de forma mais específica, a proposta focou no uso do conceito de evento compartilhado na anotação de pessoas.

Para a anotação semiautomática de pessoas, foram utilizados estimadores baseados em informações de contexto (temporais e espaciais) e de conteúdo (faces das pessoas) das fotografias. Já para a anotação automática de pessoas, foram utilizadas técnicas de detecção e reconhecimento de faces. Em seguida, para ambas as abordagens, foram detectados eventos pessoais na coleção de fotografias e, por fim, foi executado um algoritmo de detecção de eventos compartilhados, identificando conjuntos de eventos similares.

Desta forma, os objetivos específicos O1 e O2 foram alcançados com os algoritmos de anotação automática e semiautomática de pessoas utilizando o conceito de evento compartilhado, os quais foram descritos no capítulo 4.

Tanto para a anotação automática, quanto para a anotação semiautomática de pessoas, foi utilizada a técnica de validação cruzada, dividindo o conjunto de fotografias utilizado em dez partes, sendo uma parte para teste e as demais para treinamento. A métrica taxa de acerto foi utilizada para avaliar a abordagem desenvolvida. Foi obtida a taxa de acerto da anotação de pessoas utilizando eventos compartilhados e a taxa de acerto considerando apenas os eventos separadamente.

Com relação à hipótese H1, o conceito de evento compartilhado também trouxe um incremento na taxa de acerto para a anotação automática de pessoas (seção 5.3.1). Pôde ser observado que a utilização do conceito de evento compartilhado aumentou a taxa média de acerto (em cerca de 11%) e reduziu o desvio padrão dos resultados.

Já com relação à hipótese H2, foi visto na seção 5.3.2 que o conceito de evento compartilhado trouxe um incremento na taxa de acerto para a anotação semiautomática de pessoas. Quando o tamanho da lista de recomendações foi maior que 9, a taxa de acerto com a utilização do conceito de evento compartilhado foi 10% superior a apenas

utilizar eventos pessoais separadamente.

É importante ressaltar que os algoritmos propostos neste trabalho foram executados em um ambiente não controlado, pois as fotografias que compuseram a base de dados utilizada nem sempre tinham as faces das pessoas presentes iluminadas e na posição frontal. Além disso, para o algoritmo de anotação semiautomática de pessoas, não é necessário que haja faces detectadas para que a anotação ocorra, tendo em vista que além da informação das faces das pessoas, foram utilizados estimadores baseados nas informações de contexto das fotografias, propiciando uma melhor robustez em detrimento de soluções que são dependentes de faces detectadas.

Outro ponto a ser destacado é o fato de que o algoritmo proposto para anotação automática de pessoas necessita das faces das pessoas presentes, não tendo uma boa taxa de acerto para fotografias sem faces detectadas. Já o algoritmo de anotação semiautomática requer que as fotografias sejam anotadas com as informações de data e hora, localização geográfica e das pessoas presentes - neste caso, se não houverem pessoas anotadas, o algoritmo continuaria funcionando, porém com uma taxa de acerto inferior.

6.1 Contribuições

As principais contribuições deste trabalho foram:

- Proposição de uma solução para realizar anotação automática de pessoas utilizando o conceito de evento compartilhado;
- Proposição de uma solução para realização de anotação semiautomática de pessoas utilizando o conceito de evento compartilhado;
- Comparação de abordagens de técnicas de anotação de pessoas e eventos em coleções de fotografias digitais;
- Taxa de acerto do algoritmo de anotação semiautomática superior a outras soluções na literatura - a exemplo de Figueirêdo et al. (2015).

6.2 Trabalhos Futuros

Como trabalhos futuros, são sugeridos alguns tópicos para dar prosseguimento à pesquisa:

- Disponibilização de um serviço web que receberá algoritmos de detecção e reconhecimento de faces e avaliará o desempenho destes na base privada de fotografias utilizada nesta pesquisa;

- Utilização da anotação automática e semiautomática de pessoas para detectar eventos compartilhados, verificando assim se a utilização destas anotações traz algum incremento na precisão do algoritmo de detecção de eventos compartilhados;
- Utilização de outras soluções para detecção e reconhecimento de faces no processo de anotação de pessoas, a exemplo de Deep Learning;
- Utilização de outros estimadores na anotação automática de pessoas - como estimadores de contexto, por exemplo;
- Utilização de novos estimadores de contexto (ex: utilização dos contatos de um usuário em redes sociais) e de conteúdo (ex: utilização da cor das roupas) na anotação semiautomática de pessoas;
- Avaliação do impacto da granularidade dos eventos nas funções de similaridade no que concerne à detecção de eventos compartilhados.
- Testar outros algoritmos de agrupamento para detecção de eventos pessoais (ex: TraClus).
- Verificar como as soluções para anotação automática e semiautomática de pessoas se comportam diante das falhas dos detectores de face utilizados.
- Utilizar detecção de objetos repetidos e de tags contidas nas fotografias para melhorar a detecção de eventos compartilhados.

Referências

- AMES, M.; NAAMAN, M. Why We Tag: Motivations for Annotation in Mobile and Online Media. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2007. (CHI '07), p. 971 – 980. Disponível em: <<http://doi.acm.org/10.1145/1240624.1240772>>.
- ANDERSON, T. W. Anderson-Darling Tests of Goodness-of-Fit. In: *International Encyclopedia of Statistical Science*. [s.n.], 2011. p. 52 – 54. Disponível em: <http://dx.doi.org/10.1007/978-3-642-04898-2_118>.
- ANDRADE, D. et al. New Approaches for Geographic Location Propagation in Digital Photograph Collections. In: *16th International Conference on Enterprise Information Systems*. [S.l.: s.n.], 2014. p. 92 – 99.
- ANDRADE, D. O. S. de et al. Photo annotation: a survey. In: *Multimedia Tools and Applications*. [S.l.: s.n.], 2016.
- BOUSELMI, G.; FOHR, D.; ILLINA, I. Multilingual recognition of non-native speech using acoustic model transformation and pronunciation modeling. *I. J. Speech Technology*, v. 15, n. 2, p. 203 – 213, 2012. Disponível em: <<http://dx.doi.org/10.1007/s10772-012-9134-8>>.
- BRENNER, M.; IZQUIERDO, E. Social Event Detection and Retrieval in Collaborative Photo Collections. In: *Proceedings of the 2Nd ACM International Conference on Multimedia Retrieval*. ACM, 2012. (ICMR '12), p. 21:1 – 21:8. Disponível em: <<http://doi.acm.org/10.1145/2324796.2324823>>.
- BRENNER, M.; MIRZA, N.; IZQUIERDO, E. People Recognition Using Gamified Ambiguous Feedback. In: *Proceedings of the First International Workshop on Gamification for Information Retrieval*. ACM, 2014. (GamifIR '14), p. 22 – 26. Disponível em: <<http://doi.acm.org/10.1145/2594776.2594781>>.
- CHOI, J. Y. et al. Collaborative Face Recognition for Improved Face Annotation in Personal Photo Collections Shared on Online Social Networks. *IEEE Trans. Multimedia*, v. 13, n. 1, p. 14 – 28, 2011. Disponível em: <<http://dx.doi.org/10.1109/TMM.2010.2087320>>.
- CHUNG, K. The Approximate Distribution of Student's Statistic. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 17, n. 4, p. 447 – 465, 1946. ISSN 00034851. Disponível em: <<http://www.jstor.org/stable/2236084>>.
- CONCI, N. et al. Synchronization of Multi-User Event Media at MediaEval 2015: TaskDescription, Datasets, and Evaluation. In: *Working Notes Proceedings of the MediaEval 2015 Workshop, Wurzen, Germany, September 14-15, 2015*. [s.n.], 2015. Disponível em: <<http://ceur-ws.org/Vol-1436/Paper9.pdf>>.
- COORAY, S. et al. IDENTIFYING PERSON RE-OCCURRENCES FOR PERSONAL PHOTO MANAGEMENT APPLICATIONS. In: *2006 IET International Conference on Visual Information Engineering*. [S.l.: s.n.], 2006. p. 144 – 149. ISSN 0537-9989.

COORAY, S. H. *Enhancing person annotation for personal photo management using content and context based technologies*. 2008. 250 p. Tese (Computer Science) — Dublin City University.

COORAY, S. H.; O'CONNOR, N. E. Enhancing Person Annotation for Personal Photo Management Applications. In: *2009 20th International Workshop on Database and Expert Systems Application*. [S.l.: s.n.], 2009. p. 251 – 255. ISSN 1529-4188.

DAO, M. et al. Jointly Exploiting Visual and Non-visual Information for Event-related Social Media Retrieval. In: *Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval*. ACM, 2013. (ICMR '13), p. 159 – 166. Disponível em: <<http://doi.acm.org/10.1145/2461466.2461494>>.

DAVIS, M. et al. From Context to Content: Leveraging Context to Infer Media Metadata. In: *Proceedings of the 12th Annual ACM International Conference on Multimedia*. ACM, 2004. (MULTIMEDIA '04), p. 188 – 195. Disponível em: <<http://doi.acm.org/10.1145/1027527.1027572>>.

DAVIS, M. et al. Towards Context-aware Face Recognition. In: *Proceedings of the 13th Annual ACM International Conference on Multimedia*. ACM, 2005. (MULTIMEDIA '05), p. 483 – 486. Disponível em: <<http://doi.acm.org/10.1145/1101149.1101257>>.

DAVIS, M. et al. Using context and similarity for face and location identification. In: . [s.n.], 2006. v. 6061, p. 60610E – 9. Disponível em: <<http://dx.doi.org/10.1117/12.650981>>.

ESCALANTE, H. J.; MONTES-Y-GÓMEZ, M.; SUCAR, L. E. Multi-class particle swarm model selection for automatic image annotation. *Expert Syst. Appl.*, v. 39, n. 12, p. 11011 – 11021, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2012.03.023>>.

ESTER, M. et al. A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 1996. (KDD'96), p. 226 – 231. Disponível em: <<http://dl.acm.org/citation.cfm?id=3001460.3001507>>.

FENG, K. et al. In Search of Influential Event Organizers in Online Social Networks. In: *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*. ACM, 2014. (SIGMOD '14), p. 63 – 74. Disponível em: <<http://doi.acm.org/10.1145/2588555.2612173>>.

FIGUEIRÊDO, H. F. de et al. Semi-automatic photograph tagging by combining context with content-based information. *Expert Syst. Appl.*, v. 42, n. 1, p. 203 – 211, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2014.07.060>>.

FIGUEIRÊDO, H. F. de et al. Detection of Photos from the Same Event Captured by Distinct Cameras. In: *Proceedings of the 18th Brazilian Symposium on Multimedia and the Web*. ACM, 2012. (WebMedia '12), p. 51 – 58. Disponível em: <<http://doi.acm.org/10.1145/2382636.2382650>>.

GALLAGHER, A.; CHEN, T. Clothing Cosegmentation for Recognizing People. In: *Proc. CVPR*. [S.l.: s.n.], 2008.

GALLAGHER, A. C.; CHEN, T. Using Context to Recognize People in Consumer Images. *IPSN Transactions on Computer Vision and Applications*, v. 1, p. 115 – 126, 2009.

GALLAGHER, A. C. et al. Image Annotation Using Personal Calendars As Context. In: *Proceedings of the 16th ACM International Conference on Multimedia*. ACM, 2008. (MM '08), p. 681 – 684. Disponível em: <<http://doi.acm.org/10.1145/1459359.1459458>>.

GAO, X. et al. GeSoDeck: A Geo-social Event Detection and Tracking System. In: *Proceedings of the 21st ACM International Conference on Multimedia*. ACM, 2013. (MM '13), p. 471 – 472. Disponível em: <<http://doi.acm.org/10.1145/2502081.2502274>>.

HANBURY, A. A survey of methods for image annotation. *Journal of Visual Languages & Computing*, v. 19, n. 5, p. 617 – 627, 2008. ISSN 1045-926X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1045926X08000037>>.

HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression (Wiley Series in probability and statistics)*. 2. ed. Wiley-Interscience Publication, 2003. ISBN 0471356328. Disponível em: <<http://www.amazon.com/Applied-logistic-regression-probability-statistics/dp/0471356328?SubscriptionId=192BW6DQ43CK9FN0ZGG2&tag=ws&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0471356328>>.

HUANG, S. C.; JIAU, M. K.; JIAN, Y. H. Optimisation of automatic face annotation system used within a collaborative framework for online social networks. *IET Computer Vision*, v. 10, n. 5, p. 349 – 358, 2016. ISSN 1751-9632.

HULSEBOSCH, R. J.; EBBEN, P. W. G. Enhancing Face Recognition with Location Information. In: *2008 Third International Conference on Availability, Reliability and Security*. [S.l.: s.n.], 2008. p. 397 – 403.

IONESCU, B. et al. Div150Multi: A Social Image Retrieval Result Diversification Dataset with Multi-topic Queries. In: *Proceedings of the 7th International Conference on Multimedia Systems*. ACM, 2016. (MMSys '16), p. 46:1 – 46:6. Disponível em: <<http://doi.acm.org/10.1145/2910017.2910620>>.

IVASIC-KOS, M.; IPSIC, I.; RIBARIC, S. A knowledge-based multi-layered image annotation system. *Expert Syst. Appl.*, v. 42, n. 24, p. 9539 – 9553, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2015.07.068>>.

JACCARD, P. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, v. 37, p. 241 – 272, 1901.

JAMIL, N.; SA'DAN, S. A. Automated face annotation for personal photo management. In: *2014 International Conference on Computational Science and Technology (ICCST)*. [S.l.: s.n.], 2014. p. 1 – 5.

JANG, C.; YOON, T.; CHO, H. A Smart Clustering Algorithm for Photo Set Obtained from Multiple Digital Cameras. In: *Proceedings of the 2009 ACM Symposium on Applied Computing*. ACM, 2009. (SAC '09), p. 1784 – 1791. Disponível em: <<http://doi.acm.org/10.1145/1529282.1529683>>.

- JANG, C.; YOON, T.; CHO, H. G. A New Clustering Methodology for Group Photos Taken by Multiple Travelers. In: *2009 Ninth IEEE International Conference on Computer and Information Technology*. [S.l.: s.n.], 2009. v. 1, p. 118 – 123.
- KIM, H.; EL-SADDIK, A.; JUNG, J. Leveraging personal photos to inferring friendships in social network services. *Expert Syst. Appl.*, v. 39, n. 8, p. 6955 – 6966, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2012.01.022>>.
- KOLMOGOROV, A. N. *Foundations of Probability*. Berlin: [s.n.], 1933.
- LACERDA, Y. A. et al. Expandindo e Utilizando InformaçÕes De Contexto Para a Sugestao De AnotaçÕes De Fotografias Digitais. In: *Proceedings of the 14th Brazilian Symposium on Multimedia and the Web*. ACM, 2008. (WebMedia '08), p. 162 – 169. Disponível em: <<http://doi.acm.org/10.1145/1666091.1666119>>.
- LACERDA, Y. A. et al. PhotoGeo: A Self-Organizing System for Personal Photo Collections. In: *2008 Tenth IEEE International Symposium on Multimedia*. [S.l.: s.n.], 2008. p. 258 – 265.
- LEONG, C. W.; MIHALCEA, R.; HASSAN, S. Text mining for automatic image tagging. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 2010. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. [S.l.], 2010. p. 647 – 655.
- LIENHART, R.; MAYDT, J. An extended set of Haar-like features for rapid object detection. In: *Proceedings. International Conference on Image Processing*. [S.l.: s.n.], 2002. v. 1, p. I-900–I-903 – vol.1. ISSN 1522-4880.
- LIM, J.; TIAN, Q.; MULHEM, P. Home photo content modeling for personalized event-based retrieval. *IEEE MultiMedia*, v. 10, n. 4, p. 28 – 37, 2003. ISSN 1070-986X.
- LIN, D. et al. Joint People, Event, and Location Recognition in Personal Photo Collections Using Cross-domain Context. In: *Proceedings of the 11th European Conference on Computer Vision: Part i*. Springer-Verlag, 2010. (ECCV'10), p. 243 – 256. Disponível em: <<http://dl.acm.org/citation.cfm?id=1886063.1886083>>.
- MADHUMATHI, K.; THANAMANI, A. Face Annotation Using Unsupervised Label Refinement and Facial Gesture Detection using Eigenfaces Algorithm. *International Journal of Advanced Research in Computer and Communication Engineering*, v. 3, n. 9, p. 7909 – 7911, 2014.
- MANN, H. B.; WHITNEY, D. R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 18, n. 1, p. 50 – 60, 1947. ISSN 00034851. Disponível em: <<http://www.jstor.org/stable/2236101>>.
- MARKUS, N. et al. Object Detection with Pixel Intensity Comparisons Organized in Decision Trees. *ArXiv e-prints*, may 2013.
- MEZARIS, V. et al. Real-life events in multimedia: detection, representation, retrieval, and applications. *Multimedia Tools Appl.*, v. 70, n. 1, p. 1 – 6, 2014. Disponível em: <<http://dx.doi.org/10.1007/s11042-013-1426-8>>.

- MONAGHAN, F.; O'SULLIVAN, D. Leveraging Ontologies, Context and Social Networks to Automate Photo Annotation. In: *Proceedings of the Semantic and Digital Media Technologies 2Nd International Conference on Semantic Multimedia*. Springer-Verlag, 2007. (SAMT'07), p. 252 – 255. Disponível em: <<http://dl.acm.org/citation.cfm?id=1780533.1780568>>.
- NAAMAN, M. et al. Automatic organization for digital photographs with geographic coordinates. In: *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, 2004*. [S.l.: s.n.], 2004. p. 53 – 62.
- NAAMAN, M. et al. Leveraging Context to Resolve Identity in Photo Albums. In: *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*. ACM, 2005. (JCDL '05), p. 178 – 187. Disponível em: <<http://doi.acm.org/10.1145/1065385.1065430>>.
- NAKAJI, Y.; YANAI, K. Visualization of Real-World Events with Geotagged Tweet Photos. In: *2012 IEEE International Conference on Multimedia and Expo Workshops*. [S.l.: s.n.], 2012. p. 272 – 277.
- O'HARE, N.; SMEATON, A. F. Context-Aware Person Identification in Personal Photo Collections. *IEEE Trans. Multimedia*, v. 11, n. 2, p. 220 – 228, 2009. Disponível em: <<http://dx.doi.org/10.1109/TMM.2008.2009679>>.
- PANIAGUA, J. et al. Social Events and Social Ties. In: *Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval*. ACM, 2013. (ICMR '13), p. 143 – 150. Disponível em: <<http://doi.acm.org/10.1145/2461466.2461492>>.
- PHAM, T. et al. Latent Semantic Fusion Model for Image Retrieval and Annotation. In: *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*. ACM, 2007. (CIKM '07), p. 439 – 444. Disponível em: <<http://doi.acm.org/10.1145/1321440.1321503>>.
- PRESTI, L. L.; CASCIA, M. L. Concurrent photo sequence organization. *Multimedia Tools Appl.*, v. 68, n. 3, p. 777 – 803, 2014. Disponível em: <<http://dx.doi.org/10.1007/s11042-012-1079-z>>.
- PSALLIDAS, F. et al. Effective Event Identification in Social Media. *IEEE Data Eng. Bull.*, v. 36, n. 3, p. 42 – 50, 2013. Disponível em: <<http://sites.computer.org/debull/A13sept/p42.pdf>>.
- RABBATH, M.; SANDHAUS, P.; BOLL, S. Analysing Facebook Features to Support Event Detection for Photo-based Facebook Applications. In: *Proceedings of the 2Nd ACM International Conference on Multimedia Retrieval*. ACM, 2012. (ICMR '12), p. 11:1 – 11:8. Disponível em: <<http://doi.acm.org/10.1145/2324796.2324810>>.
- RODDEN, K.; WOOD, K. R. How Do People Manage Their Digital Photographs? In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2003. (CHI '03), p. 409 – 416. Disponível em: <<http://doi.acm.org/10.1145/642611.642682>>.
- RUOCCO, M.; RAMAMPIARO, H. Geo-temporal distribution of tag terms for event-related image retrieval. *Information Processing & Management*, v. 51, n. 1, p. 92 –

110, 2015. ISSN 0306-4573. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0306457314000867>>.

SADLER, D. A. et al. User-Feedback on a Feature-Rich Photo Organiser. In: *2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services*. [S.l.: s.n.], 2008. p. 215 – 218. ISSN 2158-5873.

SANSONE, E. et al. Automatic Synchronization of Multi-user Photo Galleries. *IEEE Transactions on Multimedia*, v. 19, n. 6, p. 1285 – 1298, 2017. ISSN 1520-9210.

SHIMIZU, K.; NITTA, N.; BABAGUCHI, N. Learning People Co-occurrence Relations by Using Relevance Feedback for Retrieving Group Photos. In: *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*. ACM, 2011. (ICMR '11), p. 57:1 – 57:8. Disponível em: <<http://doi.acm.org/10.1145/1991996.1992053>>.

STONE, Z.; ZICKLER, T. E.; DARRELL, T. Autotagging Facebook: Social network context improves photo annotation. In: *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2008.

TURK, M.; PENTLAND, A. Eigenfaces for Recognition. *J. Cognitive Neuroscience*, MIT Press, Cambridge, MA, USA, v. 3, n. 1, p. 71 – 86, jan 1991. ISSN 0898-929X. Disponível em: <<http://dx.doi.org/10.1162/jocn.1991.3.1.71>>.

VARSNEY, L. Identity Annotation in Photo Collections: A Survey. *Camera Culture*, v. 1, n. 1, p. 1 – 12, Março 2008.

VIANA, W. et al. PhotoMap - Automatic Spatiotemporal Annotation for Mobile Photos. In: *Proceedings of the 7th International Conference on Web and Wireless Geographical Information Systems*. Springer-Verlag, 2007. (W2GIS'07), p. 187 – 201. Disponível em: <<http://dl.acm.org/citation.cfm?id=1778528.1778542>>.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. [S.l.: s.n.], 2001. v. 1, p. 1-511-1-518 – vol.1. ISSN 1063-6919.

WANG, G. et al. Seeing People in Social Context: Recognizing People and Social Relationships. In: _____. *Computer Vision – ECCV 2010: 11th european conference on computer vision, heraklion, crete, greece, september 5-11, 2010, proceedings, part v*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 169 – 182. ISBN 978-3-642-15555-0.

YAGNIK, J.; ISLAM, A. Learning People Annotation from the Web via Consistency Learning. In: *Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval*. ACM, 2007. (MIR '07), p. 285 – 290. Disponível em: <<http://doi.acm.org/10.1145/1290082.1290121>>.

ZHANG, W.; ZHANG, T.; TRETTER, D. Clothing-based person clustering in family photos. In: *2010 IEEE International Conference on Image Processing*. [S.l.: s.n.], 2010. p. 4593 – 4596. ISSN 1522-4880.

ZIGKOLIS, C. et al. Collaborative event annotation in tagged photo collections. *Multimedia Tools Appl.*, v. 70, n. 1, p. 89 – 118, 2014. Disponível em: <<http://dx.doi.org/10.1007/s11042-012-1154-5>>.

Apêndices

APÊNDICE A – Testes estatísticos

Nesta seção serão expostos os testes estatísticos realizados para verificar se o conceito de evento compartilhado trouxe algum incremento na taxa de acerto para anotação automática e semiautomática de pessoas em fotografias (seções 5.3.1 e 5.3.2).

A.1 Anotação automática de pessoas

Para a anotação automática de pessoas, foram realizados testes de normalidade e testes para verificação das hipóteses propostas. Para o teste de normalidade, a hipótese nula usualmente considerada é que os dados vêm de uma população normal. Assim, se deseja refutar a hipótese alternativa (a qual assume que os dados não vêm de uma população normal). Caso o valor p seja maior que 0.05, então pode-se considerar que a hipótese alternativa está descartada; assumindo então que os dados vêm de uma distribuição normal. Na tabela 3, os resultados podem ser visualizados.

Tabela 3 – Resultados do teste de normalidade para os dados da anotação automática de pessoas

Dados	Valor p
Utilizando eventos compartilhados	0.2103
Sem utilizar eventos compartilhados	0.7098

Fonte: Elaborado pelo autor

Conforme pôde ser verificado, para os dois conjuntos de dados os valores p foram maiores que 0.05. Assim, assume-se que os dados vêm de uma população normal. Para verificar se o conceito de evento compartilhado trouxe algum incremento na taxa de acerto na anotação automática, utilizamos o teste t-Student. Neste caso, a hipótese nula assumida foi que as duas populações tinham médias iguais. Caso o valor p seja que 0.05, pode-se refutar a hipótese nula, considerando assim que a utilização do conceito de evento compartilhado incrementou a taxa de acerto para a anotação automática de pessoas. Como pode-se ver na tabela 4, o valor p foi inferior à significância estatística utilizada (5%). Assim, os testes estatísticos respaldam a afirmação de que a utilização de eventos compartilhados melhora a taxa de acerto para anotação automática de pessoas.

Tabela 4 – Resultados do teste de hipóteses concernente a eventos compartilhados e anotação automática de pessoas

Teste	Valor p
Verificação dos resultados utilizando eventos compartilhados serem superiores aos da não utilização deste conceito	4.982e-08

Fonte: Elaborado pelo autor

A.2 Anotação semiautomática de pessoas

Foram realizados testes de normalidade e testes de verificação de hipóteses para a anotação semiautomática de pessoas. Para o teste de normalidade, a hipótese nula usualmente considerada é que os dados vêm de uma população normal. Assim, se deseja refutar a hipótese alternativa. Caso o valor p seja maior que 0.05, então pode-se considerar que a hipótese alternativa está descartada; assumindo então que os dados vêm de uma distribuição normal. Na tabela 5, podem ser visualizados os resultados.

Tabela 5 – Resultados do teste de normalidade para os dados da anotação semiautomática de pessoas

Dados	Valor p
Utilizando eventos compartilhados	9.36e-06
Sem utilizar eventos compartilhados	2.279e-08

Fonte: Elaborado pelo autor

Para os dois conjuntos de dados, os valores p foram menores que 0.05. Assim, não há suporte para rejeitar a hipótese alternativa. Deste modo, não há como considerar que os dados vêm de uma população normal. Assim, para verificar se o conceito de evento compartilhado trouxe algum incremento na taxa de acerto na anotação semiautomática, utilizamos um teste não-paramétrico (i.e., que não requer que os dados sejam normais). A hipótese nula assumida foi que as duas populações tinham médias iguais. Caso o valor p seja que 0.05, pode-se refutar a hipótese nula, considerando assim que a utilização do conceito de evento compartilhado incrementou a taxa de acerto para a anotação semiautomática de pessoas. Como pode ser visto na tabela 6, o valor p foi inferior à significância estatística utilizada (5%). Assim, os testes estatísticos respaldam a afirmação de que a utilização de eventos compartilhados melhora a taxa de acerto para anotação semiautomática de pessoas.

Tabela 6 – Resultados do teste de hipóteses concernente a eventos compartilhados e anotação semiautomática de pessoas

Teste	Valor p
Verificação dos resultados utilizando eventos compartilhados serem superiores aos da não utilização deste conceito	0.0002408

Fonte: Elaborado pelo autor