



Universidade Federal de Pernambuco

Centro de Informática - CIn

Graduação em Engenharia da Computação

Grids irregulares adaptativos para localização
outdoor de estações móveis em ambientes urbanos

Gustavo Pimentel Bittencourt

TRABALHO DE GRADUAÇÃO

Recife, Julho de 2017



Universidade Federal de Pernambuco

Centro de Informática - CIn

Graduação em Engenharia da Computação

Grids irregulares adaptativos para localização
outdoor de estações móveis em ambientes urbanos

Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Aluno: Gustavo Pimentel Bittencourt

Orientador: Daniel Carvalho da Cunha

"Imagination will often carry us to worlds that never were. But without it we go nowhere."

Carl Sagan

Agradecimentos

A todos aqueles que direta ou indiretamente
contribuíram para realização deste trabalho.

Resumo

Localizar estações móveis em um ambiente sem fio com precisão cada vez maior ainda constitui um dos principais desafios dos sistemas de posicionamento atuais. Entre as técnicas disponíveis, aquelas baseadas em *Radio Frequency Fingerprinting* (*RF fingerprinting*) se destacam devido a boa capacidade de estimar a localização de estações móveis em ambientes *indoor* e urbanos, onde nem sempre existe uma linha direta entre o transmissor e o receptor. O objetivo da localização por *fingerprinting* é criar uma espécie de assinatura de rádio frequência capaz de caracterizar unicamente um dado ponto do espaço. Uma das características da técnica consiste na forma como a região de cobertura é modelada. Por região de cobertura subentende-se toda a área na qual deseja-se localizar uma estação móvel. Essa área é tipicamente mapeada utilizando um grande retângulo que é subdividido em células quadradas menores para formar um *grid*. Uma assinatura de rádio frequência (*fingerprint*) é associada ao centro de cada célula do *grid*. O processo de localização ocorre através de algoritmos que correlacionam novos *fingerprints* com aqueles previamente assimilados pelo modelo. A ideia é que assinaturas similares possuam uma localização similar, fornecendo assim, uma boa estimativa de localização. Os trabalhos envolvendo localização baseada em *fingerprinting* focam majoritariamente no aprimoramento dos algoritmos de correlação, sendo minoria os estudos que investigam a influência da geometria do *grid* no resultado da estimativa. Baseado nessa constatação, esse trabalho propõe um novo modelo de *grid* capaz de adaptar-se as características morfológicas da malha viária dentro de um dado perímetro urbano. Foi realizado um estudo comparativo entre os modelos de *grid* encontrados na literatura com essa nova proposta apresentada. Os resultados apontam que o emprego dessa geometria é promissora para problemas de localização que envolvem espaços urbanos densamente ocupados por edificações de morfologia e distribuição espacial arbitrárias. Para validar os resultados foi utilizada uma base de dados com 2.756 medições obtidas na cidade de Recife/PE para o sistema de telefonia móvel GSM na faixa de 1.8GHz.

Palavras-chaves: localização *outdoor*, localização GSM, *grids* irregulares, *RF fingerprinting*, *RSS fingerprinting*, *Received Signal Strength* (RSS), *Received Signal Strength Indicator* (RSSI).

Lista de Figuras

1	Exemplo de <i>RF fingerprint</i> com relação ao RSS (<i>Received Signal Strength</i>) . . .	13
2	Exemplos de <i>grids</i> e a influência da resolução	15
3	Ilustração em mapa e foto de satélite da região escolhida	18
4	Localização das estações base na região escolhida	19
5	Conjunto de medições presentes na base de dados	20
6	Exemplo de <i>grid</i> regular para região das medições	21
7	Vias representadas como segmentos delimitados por coordenadas GPS	22
8	Influência do tráfego e do erro dos instrumentos na coleta de <i>RF fingerprints</i> . .	23
9	Conjunto de medições presentes na base de dados	24
10	Coordenadas GPS obtidas utilizando as funcionalidades da Google Roads API .	25
11	Exemplo de <i>grid</i> irregular com resolução de <i>3 metros</i>	25
12	Conjunto de dados para treinamento e teste dos <i>grids</i>	26
13	Diagrama em blocos para o sistema de localização	27
14	Influência da <i>resolução</i> na quantidade de células presentes no <i>grid</i>	30
15	Relação entre a resolução dos <i>grids</i> e a quantidade de células	31
16	Tempo de duração associado a etapa de treinamento dos <i>grids</i>	33
17	Relação entre o tempo de treinamento e a resolução dos <i>grids</i>	33
18	Tempo de duração associado a etapa de predição dos <i>grids</i>	35
19	Ajuste de curva para o tempo de predição com o número de células no <i>grid</i> . . .	35
20	Relação entre o tempo de predição e a resolução dos <i>grids</i>	36
21	Ajuste de curva para o tempo de predição com o parâmetro <i>resolução</i>	36
22	Consumo de memória volátil com o aumento do número de células	38
23	Ajuste de curva para o consumo de memória volátil com o número de células . .	38
24	Relação entre a memória para treinamento e a resolução dos <i>grids</i>	39
25	Ajuste de curva para o consumo de memória volátil com a resolução do <i>grid</i> . .	40
26	Ajuste de curva para o consumo de memória volátil com o número de células . .	41
27	Relação entre a memória para teste e a resolução dos <i>grids</i>	41
28	Comparativo entre os modelos de <i>grid</i> quanto a precisão da estimativa	44
29	Tendência de diminuição do erro com relação a <i>resolução</i>	45

Lista de Tabelas

1	Relação das estações base presentes na base de dados	21
2	Quantidade de células em cada modelo de <i>grid</i> conforme a <i>resolução</i>	31
3	Erro associado a estimativa de localização para os modelos de <i>grid</i>	44

Sumário

1	Prólogo	8
1.1	Introdução	8
1.2	Motivação e Objetivos	9
1.3	Estrutura do Documento	10
2	Contextualização	11
2.1	Técnicas e Abordagens	11
2.2	RF Fingerprinting	12
2.3	RSS e RSSI	13
2.4	Grids	14
2.5	Radio Maps	15
3	Metodologia	17
3.1	Escolha da Região	17
3.2	Criação dos <i>Grids</i>	21
3.2.1	<i>Grids</i> Regulares	21
3.2.2	<i>Grids</i> Irregulares	22
3.3	Algoritmo Proposto	26
4	Resultados	29
4.1	Tamanho dos <i>Grids</i>	29
4.2	Tempo de Duração	32
4.2.1	Etapa de Treinamento (<i>offline</i>)	32
4.2.2	Etapa de Predição (<i>online</i>)	34
4.3	Consumo de Memória Volátil	37
4.3.1	Etapa de Treinamento (<i>offline</i>)	38
4.3.2	Etapa de Predição (<i>online</i>)	40
4.4	Erro na Estimativa	42
5	Conclusão	47

1 Prólogo

Com a finalidade de facilitar a compreensão esse Capítulo foi dividido em três partes. A primeira parte introduz de forma breve o problema de localização de estações móveis. Em seguida é apresentada a principal motivação e os objetivos propostos para o desenvolvimento desse estudo. A última parte encerra apresentando uma descrição geral sobre a estrutura desse documento com o propósito de facilitar sua compreensão e leitura.

1.1 Introdução

Localizar estações móveis em um ambiente sem fio com precisão cada vez maior ainda constitui um dos principais desafios dos sistemas de posicionamento atuais. Ao longo das últimas décadas essa necessidade impulsionou o desenvolvimento de um conjunto de técnicas capazes de estimar a localização de pessoas e objetos nos mais diversos cenários [1, 2, 3].

Os serviços de localização são conhecidos na literatura pela sigla LBS (*Location Based Services*). As aplicações que fazem uso desse tipo de tecnologia podem variar em contexto e finalidade. Entre os exemplos mais clássicos podemos destacar aquelas aplicações que informam a respeito das condições do clima ou do tráfego com base na localização atual do usuário. Existem ainda os casos de uso mais complexos, como aqueles aplicáveis no contexto de detecção e prevenção de fraudes. Nesse cenário é possível correlacionar, por exemplo, o posicionamento do usuário no momento da utilização do cartão de crédito com a localização do estabelecimento comercial para o qual a transação foi efetuada.

Entre as técnicas disponíveis, aquelas baseadas em *Radio Frequency Fingerprinting*, ou simplesmente *RF fingerprinting*, se destacam devido a capacidade de estimar a localização de estações móveis mesmo com eventuais obstruções na linha de visada (*Line Of Sight* ou LOS). Essa característica tornam tais técnicas bastante atrativas para o emprego em problemas que envolvem a localização em ambientes *indoor* e urbanos, onde nem sempre existe uma linha direta entre o transmissor e o receptor [4].

Em linhas gerais um *RF fingerprint* consiste em um conjunto de características observáveis e/ou previstas a respeito do sinal de rádio interceptado em um dado ponto do espaço. O conjunto de características é utilizado como uma espécie de assinatura capaz de mapear unicamente esse ponto. Essas assinaturas são combinadas juntamente com a informação de localização espacial para construir uma base de dados que encontra-se indicada ao longo desse documento

através do termo *reference DB*. Os *RF fingerprints* associados as coordenadas espaciais e armazenados no *reference DB* são chamados de *reference RF fingerprints*.

A informação de localização espacial presente nos registros do *reference DB* corresponde a uma determinada região de interesse na qual deseja-se encontrar o posicionamento de estações móveis. Essa região é comumente modelada como um grande retângulo que é subdividido em um conjunto de células quadradas menores para formar um *grid* [5]. O comprimento de cada célula é fixado por um parâmetro chamado *resolução* que é tipicamente definido em unidades de metro. O centro de cada célula aferido em um dado sistema de coordenadas corresponde exatamente a informação localização espacial de cada registro.

Para as técnicas de localização fundamentadas nas informações de *RF fingerprinting* é comum encontrar casos de uso onde os dados armazenados no *reference DB* são utilizados como insumo para modelos estatísticos ou baseados em algoritmos de aprendizagem de máquina [1, 2]. Esses algoritmos são utilizados com o intuito de melhor adaptar as características que compõem os *RF fingerprints* às coordenadas de cada célula presente no *grid*. Essa adaptação ocorre tipicamente empregando fases de treinamento utilizando o conjunto de dados que integra o *reference DB*.

Após etapa de treinamento, o *grid* passa a constituir o que chamada-se de *Correlation Database* (CDB) ou simplesmente *radio map*. Essa nova base de dados por sua vez é utilizada como referência para a etapa de predição, onde procura-se correlacionar as assinaturas de RF nele presentes com o valor obtido para novas medições de RF apresentadas ao modelo.

Baseando-se nessa construção, o problema de localizar uma estação móvel torna-se equivalente ao problema de determinar o grau de similaridade entre a assinatura obtida em uma dada estação móvel, chamada de *target RF fingerprint*, com aquelas presentes no *Correlation Database*. Assinaturas similares devem fornecer uma boa estimativa a respeito da localização espacial da estação móvel em questão.

1.2 Motivação e Objetivos

A literatura a respeito da utilização e do aprimoramento de técnicas de aprendizagem de máquina para construção dos banco de dado para correlações é bastante extensa [1, 2, 6]. Apesar disso, pouca investigação existe a respeito do impacto da escolha da geometria do *grid* na acurácia dessas técnicas [7, 8].

Sendo assim, o objetivo deste trabalho consiste em investigar o impacto da utilização de

grids irregulares na acurácia das técnicas de localização em ambientes urbanos densamente ocupados. Tais ambientes são caracterizados por possuírem edificações diversas, com variações arbitrárias quanto a sua morfologia e distribuição espacial.

Nos *grids* regulares a diminuição do parâmetro *resolução* leva a uma conseqüente melhora na precisão da estimativa de localização. Entretanto, a eficácia desse ganho tende diminuir à medida que a *resolução* se torna cada vez menor [9, 10]. A existência dessa relação também será investigada para os *grids* irregulares.

Para condução deste estudo foi escolhida uma área urbana cujas medições de RF foram realizadas para o sistema de telefonia móvel GSM (*Global Mobile System Communication*) com portadora na faixa de frequência de 1.8 GHz. A base de dados conta com 2.756 medições obtidas na cidade de Recife/PE [6].

1.3 Estrutura do Documento

Este trabalho encontra-se organizado da seguinte maneira: o Capítulo 2 apresenta uma breve contextualização a respeito do problema de localização, definindo conceitos importantes que serão utilizados no restante desse documento. No Capítulo 3 é detalhada a metodologia adotada para condução desse estudo, o conjunto de dados utilizado para validação é apresentado juntamente com o processo para construção dos *grids* e o cenário de execução dos experimentos. O Capítulo 4 debate a respeito dos resultados obtidos, fazendo um comparativo entre ambos os modelos de *grid* escolhidos. Por fim, o Capítulo 5 apresenta uma conclusão delineando alguns pontos de possíveis questionamentos a respeito das técnicas de localização baseadas no modelo de *grid* irregular proposto nesse estudo.

2 Contextualização

2.1 Técnicas e Abordagens

São inúmeras as técnicas disponíveis atualmente para localização de estações móveis em ambientes sem fio. Todas essas técnicas tem como premissa fundamental inferir a localização de uma dada estação móvel com base em uma ou mais características mensuráveis a respeito do sinal oriundo dessa estação. Tais características são normalmente coletadas em pontos específicos cujas localizações são previamente conhecidas. Tais pontos podem corresponder, por exemplo, a um conjunto de estações base (*BSeS* ou *Base Stations*) em um sistema de telefonia GSM ou a pontos de acesso (*APs* ou *Access Points*) em uma rede Wi-Fi. Entre algumas das características mensuráveis no sinal recebido estão, por exemplo, o tempo de viagem, angulação de chegada, frequência de transmissão, atenuação de potência com relação ao sinal transmitido, entre outras.

Quanto aos algoritmos empregados, podemos classificar as técnicas de localização em três grupos: as que derivam uma interpretação geométrica com base nas informações coletadas, aquelas baseadas em modelos de propagação eletromagnética e, por último, as fundamentadas em algoritmos de aprendizado de máquina.

O primeiro grupo de técnicas está conceitualmente baseado em modelos matemáticos que tentam criar uma abstração geométrica para fornecer uma estimativa de localização com base em uma ou mais características mensuradas a respeito do sinal recebido. Entre as principais técnicas apoiadas nesse modelo estão as de multilateração e multiangulação. Essas técnicas procuram fornecer estimativas de localização com base, respectivamente, na distância percorrida e no ângulo de chegada do sinal. Para seu correto funcionamento, o sinal oriundo da estação móvel deve ser interceptado em múltiplos pontos cuja posição é previamente conhecida. Entre as técnicas mais tradicionais que podem ser classificadas segundo essa abordagem estão as baseadas em Tempo de Chegada / *Time Of Arrival* (TOA), Diferença entre Tempos de Chegada / *Time Difference Of Arrival* (TDOA) e Ângulo de Chegada / *Angle Of Arrival* (AOA).

No segundo grupo podemos enquadrar os modelos de propagação eletromagnética que tentam prever a perda com relação à potência de transmissão com base na distância percorrida pelo sinal. Sendo assim, conhecendo a potência de transmissão e medindo a potência recebida em um dado ponto do espaço é possível estimar a distância percorrida pelo sinal. Tais modelos

são baseados em formulações puramente matemáticas como o modelo para propagação em espaço livre (*Free Space Propagation Model*) [11], ou em modelos híbridos calibrados a partir de dados obtidos empiricamente como ocorre com o *Okumura-Hata* [12], *COST-231* [13], *ECC-33* [14] e *SUI (Stanford University Interim)* [15].

O terceiro grupo de técnicas apoia-se em modelos estatísticos ou em algoritmos de aprendizado de máquina. Essas técnicas objetivam minimizar o erro associado ao processo de regressão ou aferir a similaridade entre uma nova medição com aquelas previamente assimiladas pelo modelo. Essa assimilação prévia envolve uma etapa de treinamento e calibração do algoritmo utilizando uma base de dados de medições de campo obtidas empiricamente.

A fase de medições e consequente construção da base é focada na aquisição de conjunto de características do sinal recebido para cada ponto do espaço dentro de uma dada região, seja um perímetro urbano ou um espaço interno. Sendo assim, cada medição obtida nessa etapa é associada a um ponto do espaço, ponto esse que corresponde tipicamente as coordenadas GPS do local onde a medição foi efetuada. Esse conjunto de medições é armazenado e posteriormente utilizado como insumo para as etapas de treinamento dos algoritmos. Entre as abordagens baseadas nessas técnicas podemos mencionar a utilização de algoritmos como Máquina de Vetores de Suporte / *Support Vector Machine* (SVM), k-Vizinhos mais Próximos / *k Nearest Neighbours* (kNN), Redes Neurais Artificiais / *Artificial Neural Networks* (ANN), entre outras possibilidades.

2.2 RF Fingerprinting

Em linhas gerais um *RF fingerprint (Radio Frequency Fingerprint)*, consiste em um conjunto de características (vetor de características) mensuráveis a respeito do recebimento de sinais de rádio em um dado ponto do espaço. O conjunto de características é utilizado como uma espécie de assinatura capaz de mapear unicamente esse ponto. Essas assinaturas são combinadas juntamente com a informação de localização espacial para criar uma base de dados para futuras consultas conhecida como *reference DB*.

A Figura 1 a seguir ilustra um exemplo de um *RF fingerprinting* com algumas das características observáveis a respeito do sinal emitido por uma estação móvel. Perceba que o mesmo conjunto de medições se repete para cada estação base (BS) capaz de escutar o sinal oriundo da estação móvel.



Figura 1: Exemplo de *RF fingerprint* com relação ao RSS (*Received Signal Strength*)

As técnicas de localização que empregam *RF fingerprinting* pertencem a classe daquelas que envolvem modelos estatísticos ou de aprendizagem para estimar a localização de uma dada estação móvel. As técnicas que empregam *RF fingerprinting* se destacam devido a capacidade de estimar a localização de estações móveis mesmo com eventuais obstruções na linha de visada (LOS). Essa característica tornam tais técnicas bastante atrativas para o emprego em problemas que envolvem a localização em ambientes *indoor* e urbanos, onde nem sempre existe uma linha direta entre o transmissor e o receptor [4].

2.3 RSS e RSSI

Conforme indicado na seção anterior, são inúmeras as características mensuráveis que podem ser utilizadas para construção de um *RF fingerprint*. Apesar dessa aparente diversidade, a obtenção de algumas dessas medições exige equipamento especializado que normalmente não está disponível para utilização prática. Limitações no *hardware* dos dispositivos como ocorre, por exemplo, com aparelhos de telefonia celular, impede que algumas dessas medições sejam efetuadas.

Em termos práticos, as medições que compõem um *RF fingerprint* limitam-se a um pequeno conjunto composto muitas vezes de apenas uma única medição. Alguns protocolos de acesso ao meio possibilitam extrair indicadores a respeito da camada física. Entre esses indicadores, uma unanimidade presente na grande maioria dos protocolos de acesso ao meio para ambientes sem fio é o *Received Signal Strength Indicator* (RSSI).

O RSSI consiste em um parâmetro tipicamente encontrado em protocolos de camada física para indicar a qualidade do sinal recebido. Esse indicador é construído a partir da quantização do *Received Signal Strength*, ou simplesmente RSS. O RSS consiste em uma média a respeito da potência ou energia captada por um receptor. Para conseguir distinguir um sinal transmitido do ruído de fundo, o valor do RSS tem que estar acima de um limiar definido pelo projetista do rádio receptor.

Protocolos modernos de acesso ao meio permitem, por exemplo, utilizar o valor do RSSI para tomar decisões como, por exemplo, solicitar ao emissor um aumento de potência de transmissão

no intuito de aumentar a taxa de recebido ou sua diminuição com o objetivo de poupar energia. Tais protocolos são classificados na literatura como adaptativos, podendo solicitar tanto o aumento de potência como sua diminuição com base em heurísticas predefinidas.

A popularidade do RSS ou RSSI para utilização em *RF fingerprints* está ligada basicamente a dois fatores. O primeiro está associado a facilidade de obtenção dessa medida, já que seu valor é normalmente disponibilizado e encontra-se acessível para consulta na camada L1, e em alguns casos até mesmo na camada L2, dos protocolos de comunicação. O segundo fator está ligado ao bom desempenho que essas medidas possuem na capacidade de estimar o posicionamento de estações móveis mesmo com obstrução da linha de visada (*Non-Line Of Sight* ou NLOS).

Apenas exemplificando, os protocolos utilizados em redes Wi-Fi (802.11 a/b/g/n) e GSM possibilitam obter o valor para o RSSI. Entretanto, a depender da tecnologia, outros indicadores podem ser obtidos. Nos sistemas GSM o valor do RTD (*Real Time Difference*) pode ser consultado para a melhor célula, aquela ao qual o dispositivo móvel está associado [16, 17].

Por fim, durante a leitura desse documento é comum verificar que os termos *RF fingerprint* e *RSS fingerprint* são constantemente intercambiados. O termo *RF fingerprint* está ligado ao conceito mais abrangente, conforme apresentado na Seção 2.2. Já o termo *RSS fingerprint* está relacionado a um *RF fingerprint* construído considerando exclusivamente medições de RSS ou RSSI.

2.4 Grids

O cerne do estudo das técnicas de localização está em determinar o posicionamento de uma estação móvel em uma dada região de interesse ou área de cobertura. Essa região pode corresponder, por exemplo, ao interior de uma edificação ou a um dado perímetro urbano, sendo comumente modelada através de um retângulo que compreende toda sua área.

Como discutido anteriormente, é necessário estabelecer uma característica idealmente única associada a recepção do sinal de rádio em locais compreendidos dentro dessa região de cobertura. Sendo assim, é comum dividir esse retângulo em células menores que constituem pequenos quadrados cujo comprimento do lado é definido por um parâmetro chamado *resolução*, sendo esse último, tipicamente definido em unidades de metro. Ao final dessa etapa, o retângulo original transforma-se em um conjunto de células menores conhecido como *grid*.

A Figura 2 (a) mostra um *grid* estabelecido com o propósito de mapear uma dada região. O parâmetro *resolução* corresponde ao tamanho do lado associado a cada célula presente na

imagem. Vale ressaltar que tal parâmetro influencia diretamente na quantidade de células presentes no *grid*. Na Figura 2 (b) o parâmetro *resolução* foi definido como a metade do valor estipulado na Figura 2 (a).

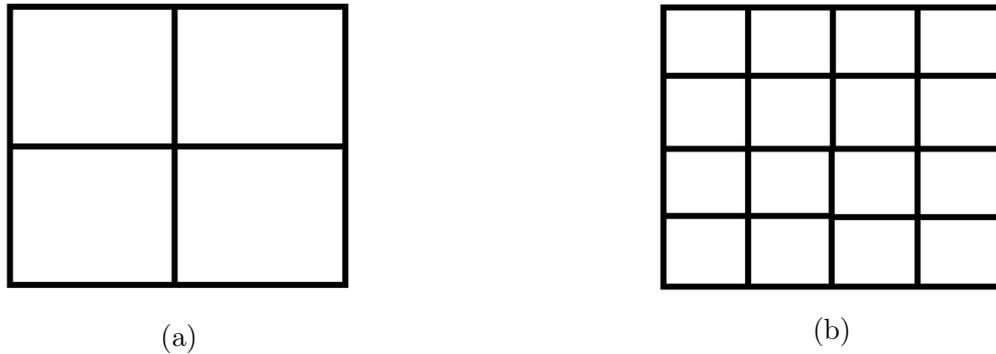


Figura 2: Exemplos de *grids* e a influência da resolução

Ainda sobre a Figura 2, perceba que o centro de cada uma das células corresponde exatamente a um ponto do espaço. Como o *grid* pode ser sobreposto em uma região real, o centro das células deve possuir latitude e longitude bem definidas através do sistema de posicionamento global (GPS - *Global Positioning System*).

Parte fundamental das técnicas de localização, principalmente aquelas fundamentadas em modelos estatísticos ou algoritmos de aprendizado de máquina, consiste em estabelecer mecanismos que possibilitem efetuar o enriquecimento desse *grid* com informações de *RF fingerprints* para cada célula. O objeto resultante dessa etapa de enriquecimento também é conhecido como *Radio Map*. Mais detalhes sobre esses objetos serão apresentados na Seção 2.5.

2.5 Radio Maps

As técnicas baseadas em *RF fingerprinting* estão apoiadas na premissa de que flutuações na potência do sinal recebido por um dado dispositivo ao longo do espaço podem ser utilizadas para montar uma base de dados para futuras correlações conhecida como *Radio Map*. Sendo assim, eventuais medições oriundas de uma dada estação móvel podem ser comparadas com aquelas presentes nessa base de medições para fornecer uma estimativa da sua localização.

Entretanto, o processo de aquisição de *RF fingerprints* possui natureza empírica. Em regiões urbanas, essa aquisição de dados ocorre tipicamente definindo uma rota que deve ser percorrida utilizando um automóvel capaz de carregar o equipamento necessário para aferir as medições.

Devido a essa característica prática é comum finalizar o processo de coleta com medições

cujas coordenadas espaciais são muito próximas entre si. Fato esse que ocorre frequentemente no caso da presença de semáforos ou congestionamentos, além de fatores inerentes ao percurso como ruas sem saída, retornos, desvios, entre outros. Outra possibilidade bastante razoável é que haja regiões inacessíveis, principalmente quando consideramos um cenário urbano. Tais regiões podem ser caracterizadas, por exemplo, por propriedades privadas ou áreas com relevo desfavorável para o tipo de veículo utilizado na obtenção das medições.

Para tornar a construção do *Radio Map* possível o espaço (2D ou 3D) é discretizado em pontos igualmente espaçados para formar o que é denominado de *grid*. A distância de espaçamento entre os pontos em um *grid* é chamado *resolução*. A cada ponto do *grid* é atribuído uma coordenada conhecida em um dado sistema de posicionamento como por exemplo o GPS (*Global Positioning System*).

A atribuição de *reference RF fingerprints* a cada um desses pontos no *grid* pode ser feita através de modelos matemáticos de propagação que preveem a perda na potência do sinal com a distância em relação ao transmissor, através de coletas de campo obtidas utilizando equipamento específico ou utilizando de uma combinação dessas duas modalidades (abordagem híbrida).

O *Radio Map* pode então ser sumarizado como um conjunto de pontos em um dado sistema de posicionamento que descreve uma área ou uma determinada região. Para cada um desses pontos é associado um conjunto de medições relacionadas a potência do sinal recebido de transmissores cuja localização e potência de transmissão são conhecidas. Essas medições descrevem um padrão que idealmente será único para cada ponto considerado no *Radio Map*.

Existem uma série de métodos para avaliar numericamente a similaridade entre dois *RF fingerprints*, seja através de métricas ou coeficientes de correlação tais como aqueles propostos pelos métodos de Euclides, Mahalanobis, Manhattan, entre outros, ou seja através de técnicas de aprendizagem de máquina.

Tópicos envolvendo o aprimoramento de algoritmos para correlação de *RF fingerprints* nos mais diversos cenários continuam dominando os principais tópicos de pesquisa na área de localização móvel [1, 2].

Para fins de esclarecimento o termo *reference RF fingerprint* está associado aos *RF fingerprints* obtidos empiricamente através de medições em campo e separados para futuro processamento. O termo também é utilizado para se referir aos *RF fingerprints* associados as coordenadas de um *Radio Map*. Já o termo *target RF fingerprint* está relacionado ao padrão obtido de uma dada estação móvel cuja localização deseja-se estimar.

3 Metodologia

Conforme mencionado anteriormente, o cerne desse trabalho consiste em verificar se existe alguma diminuição do erro associado as estimativas de localização quando utiliza-se *grids* irregulares ao invés dos *grids* tradicionalmente empregados na literatura. Esses últimos serão por brevidade denominados de *grids* regulares para o restante desse documento.

A hipótese a ser verificada conjectura que em cenários urbanos e densamente ocupados por edificações de morfologia e distribuição arbitrárias, o problema de localização *outdoor* pode ser reduzido para um problema de identificar em qual região da malha viária urbana uma dada estação móvel encontra-se.

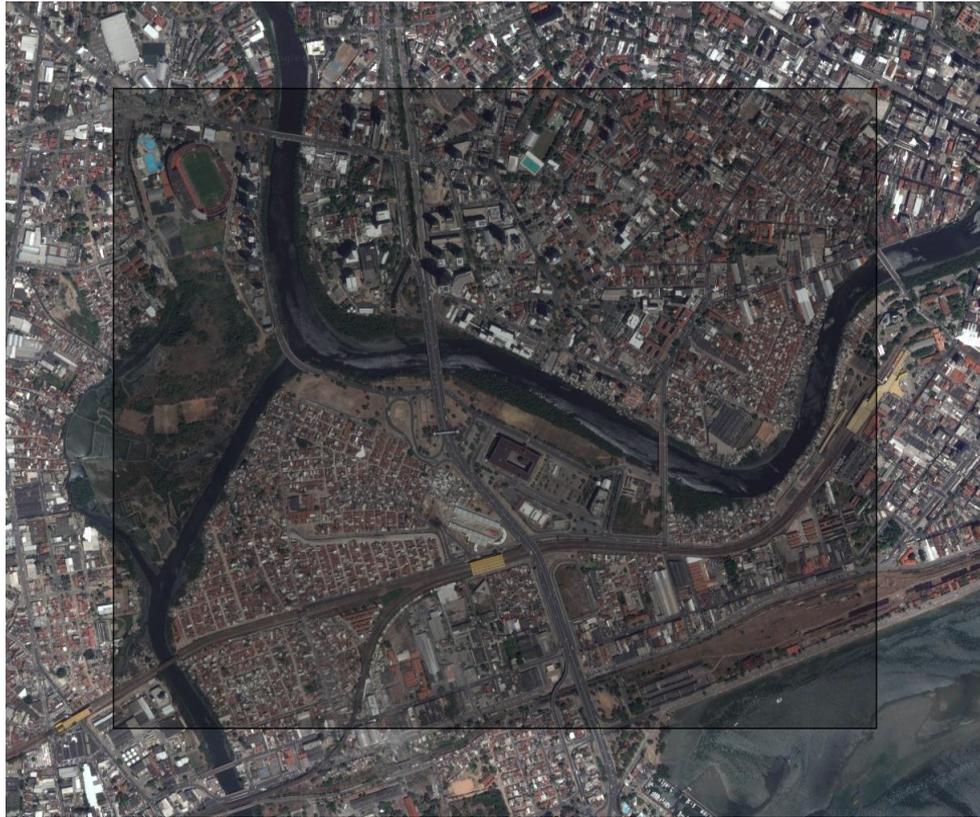
Nesse contexto, um *grid* irregular poderia ser concebido de forma customizada para adaptar-se a malha viária urbana, melhorando assim a precisão da estimativa fornecida através dos algoritmos de localização.

3.1 Escolha da Região

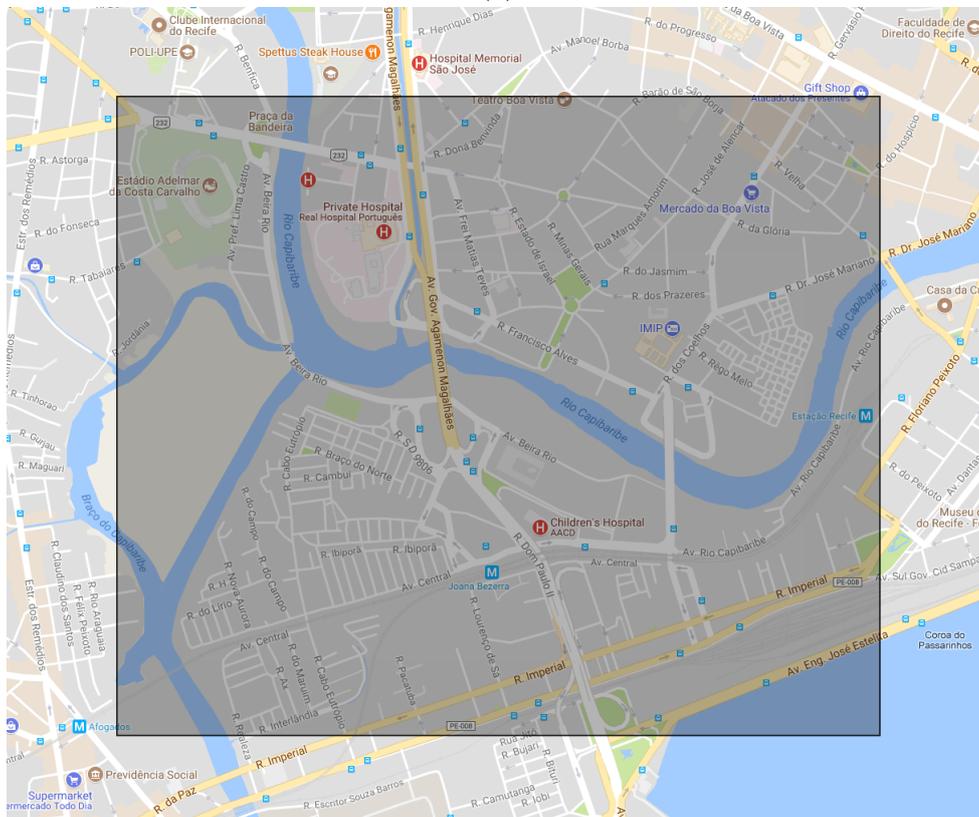
Para condução desse estudo e conseqüente validação da hipótese foi escolhida uma área urbana que corresponde a uma região central na cidade de Recife/PE. A escolha dessa região ocorreu majoritariamente devido a dois argumentos principais. Primeiramente a região atende as características exigidas para validação da hipótese no que diz respeito a densidade e morfologia das edificações, além dos requisitos de distribuição espacial. Outro fator considerado na escolha foi a existência de uma base de dados que conta com 2.756 medições para o sistema de telefonia móvel GSM (*Global Mobile System Communication*) com portadora na faixa de frequência de 1.8 GHz [6].

A região em questão encontra-se detalhada na Figura 3 sendo delimitada pelo retângulo cujos vértices correspondem as seguintes coordenadas GPS: $(-8.060549, -34.905396)$, $(-8.060549, -34.885067)$, $(-8.077546, -34.905396)$, $(-8.077546, -34.885067)$. As coordenadas estão representadas no formato (*Latitude, Longitude*).

Acima da Figura 3 conforme visualizado em (a) é apresentada uma imagem de satélite da região. Já abaixo como mostrado em (b) é exibido um mapa destacando de forma mais adequada as ruas e os trechos destinados as áreas ocupação imobiliária. Através da imagem de satélite é possível vislumbrar que as edificações e sua distribuição atendem as especificações sugeridas para condução desse estudo.

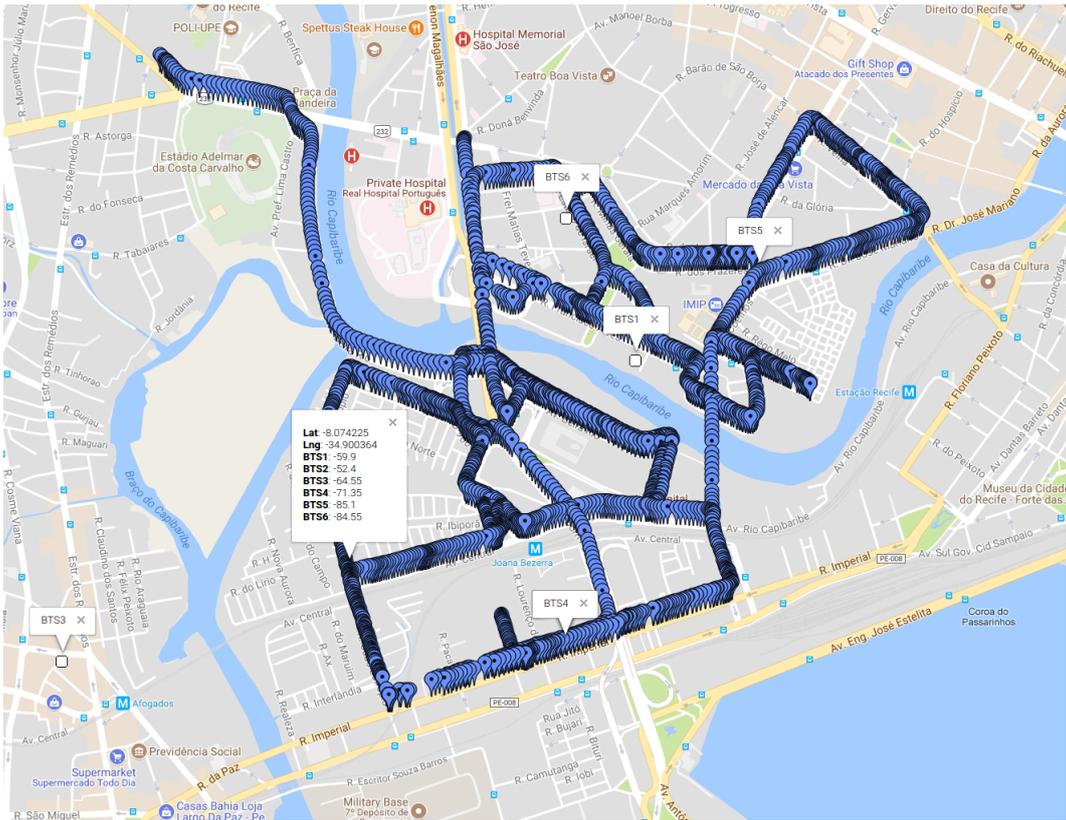


(a)



(b)

Figura 3: Ilustração em mapa e foto de satélite da região escolhida



(a)



(b)

Figura 5: Conjunto de medições presentes na base de dados

Nome	Latitude	Longitude	Freqüência	BBCH	EIRP
BTS1	-8.068361111	-34.892722222	1800 Hz	662 Hz	55.59 dBm
BTS2	-8.075916667	-34.894611111	1800 Hz	806 Hz	55.59 dBm
BTS3	-8.076361111	-34.908000000	1800 Hz	666 Hz	55.59 dBm
BTS4	-8.075916667	-34.894611112	1800 Hz	690 Hz	55.59 dBm
BTS5	-8.066000000	-34.889444444	1800 Hz	673 Hz	55.59 dBm
BTS6	-8.064583333	-34.894583333	1800 Hz	682 Hz	55.59 dBm

Tabela 1: Relação das estações base presentes na base de dados

3.2 Criação dos *Grids*

Com base na região escolhida ambos os modelos de *grid*, regular e irregular, foram construídos para mapear a área onde as medições foram efetuadas.

3.2.1 *Grids* Regulares

O *grid* regular foi construído observando as características encontradas na literatura [5], sendo assim, a região escolhida foi modelada através de um grande retângulo cujas arestas correspondem as coordenadas GPS na fronteira que compreende a região de interesse. Esse retângulo foi subdividido posteriormente em células menores com base no valor do parâmetro definido para *resolução* do *grid*.

A figura 6 ilustra um exemplo de *grid* regular montado com base na área onde as medições foram realizadas. O parâmetro *resolução* é responsável por definir o comprimento do lado associado a cada célula quadrada mostrada na imagem. A coordenada GPS atrelada ao centro de cada uma das células é considerada como um registro único. O conjunto desses registros é utilizado para criar a base de dados que representa o *grid*.

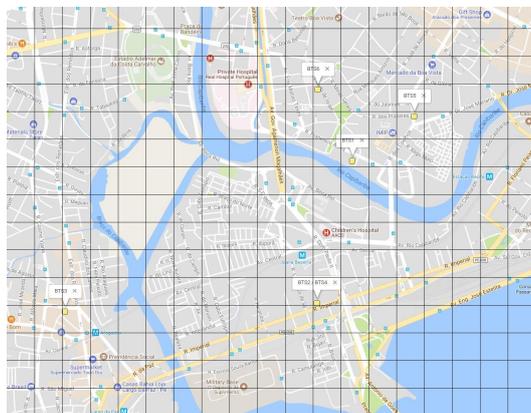
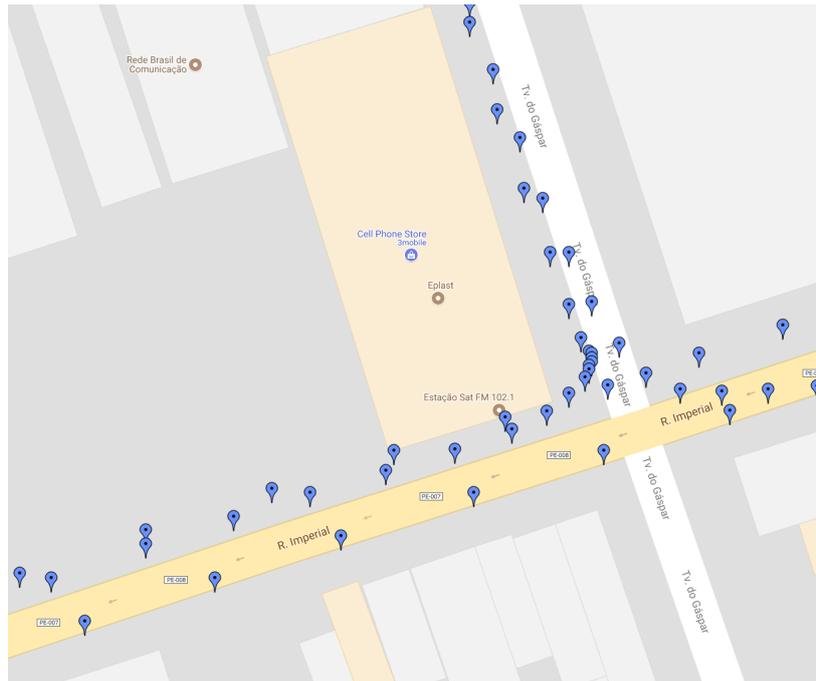
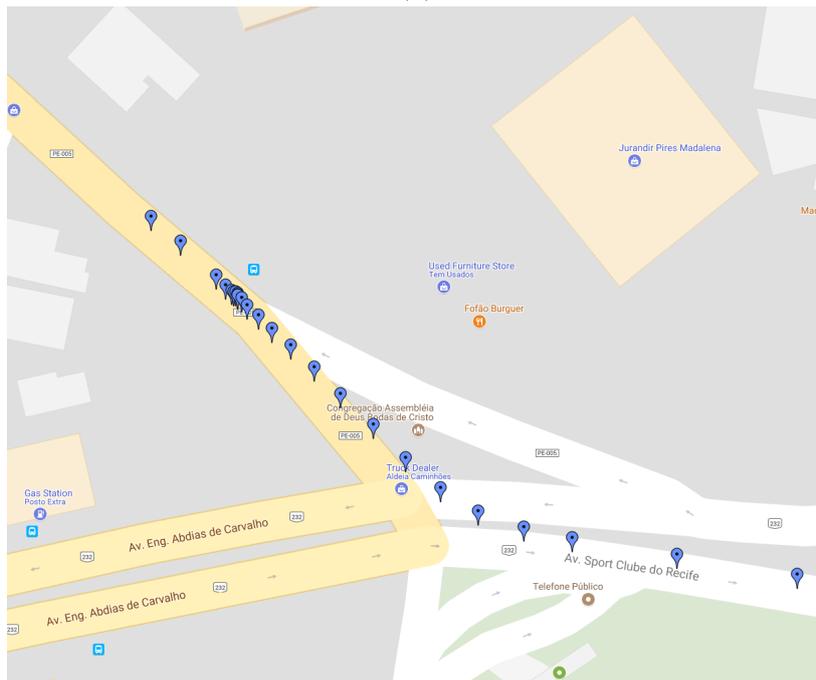


Figura 6: Exemplo de *grid* regular para região das medições



(a)



(b)

Figura 8: Influência do tráfego e do erro dos instrumentos na coleta de *RF fingerprints*

Apesar dos dados oriundos do GPS indicarem uma trajetória ao longo de uma via, sua utilização para construção dos segmentos não é tão imediata. As coordenadas GPS obtidas através do equipamento de medição possuem erros inerentes ao processo de coleta. Tais erros acabam por deslocar as coordenadas para fora da região que corresponde a via conforme indicado na Figura 8 (a). Além disso, o processo de medição no cenário avaliado está sujeito as adversi-

dades do tráfego de veículos e pedestres das grandes cidades, acarretando em lentidões devido a retenções no fluxo ou a presença de semáforos. Esses eventos fazem com que as coordenadas se acumulem nos arredores de um ou mais pontos da via como delineado na Figura 8 (b).

Para mitigar os efeitos indicados e permitir a construção do *grid* conforme idealizado, foi utilizada a Google Roads API para obtenção das coordenadas que representam os segmentos [18]. Essa API possui um recurso conhecido como *snap to roads*. Esse recurso permite ajustar uma sequência de coordenadas GPS para uma rota conforme indicado na Figura 9.

Na Figura 9 (a) temos as coordenadas GPS obtidas durante a fase de coleta, essas coordenadas possuem um erro de medição inerente ao instrumento utilizado. Na Figura 9 (b) fica evidente o resultado do ajuste realizado para os pontos indicados na Figura 9 (a). Outro recurso disponibilizado pela API permite que as coordenadas GPS sejam ajustadas para a via e interpoladas para melhor representar seu contorno. A Figura 9 (c) mostra o resultado obtido através da utilização desse recurso com relação as coordenadas indicadas na Figura 9 (a).



Figura 9: Conjunto de medições presentes na base de dados

As técnicas disponibilizadas através da Google Roads API foram aplicadas ao conjunto de coordenadas GPS extraídas do *reference DB*, resultando nos pontos ilustrados a partir da Figura 10 (a). Para efeitos de comparação, a base de medições contava originalmente com 2.756 coordenadas enquanto o conjunto de pontos que representa a malha viária para região possui apenas 347 coordenadas GPS.

Os segmentos obtidos através do conjunto com as 347 coordenadas foram interpolados utilizando a funcionalidade da Google Roads API com o objetivo de melhor ajustar os pontos aos contornos da malha viária. O conjunto originalmente com 347 coordenadas GPS foi expandido para 1.572 pontos conforme indicado na Figura 10 (b).

Por fim, os segmentos definidos pelas 1.572 coordenadas GPS foram intercalados com mais pontos com base no valor estabelecido para o parâmetro *resolução*. A Figura 11 mostra o

resultado final obtido ao intercalar as 1.572 coordenadas utilizando uma resolução de *3 metros*. A utilização desse valor para o parâmetro *resolução* implica que a maior distância possível entre dois pontos que compõem o *grid* é 3 metros.

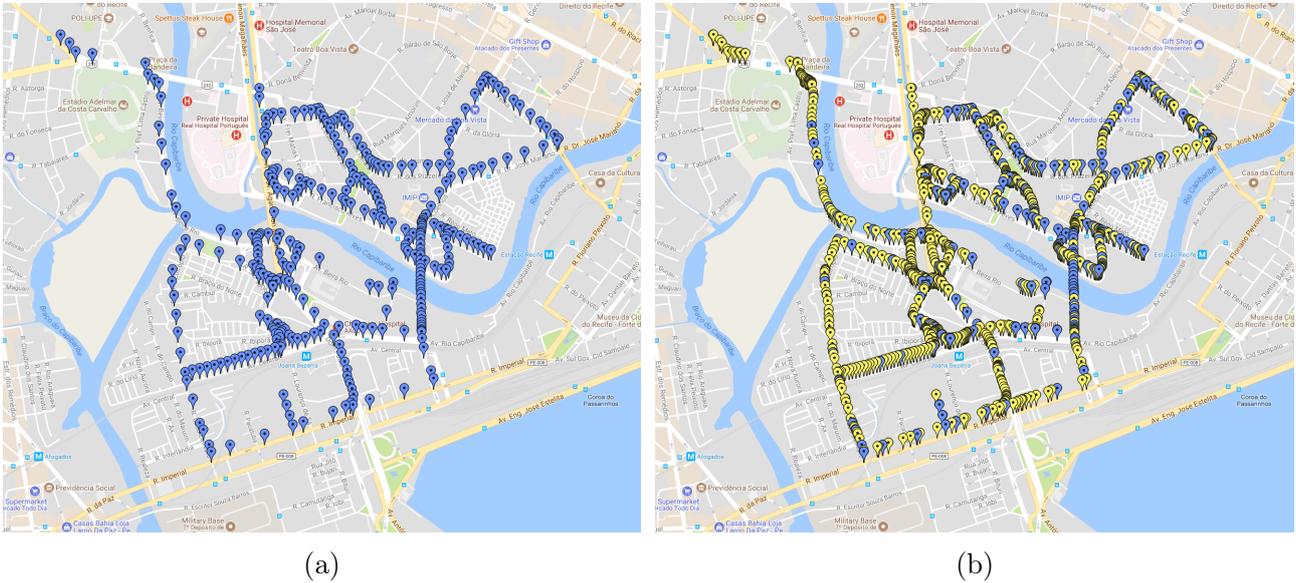


Figura 10: Coordenadas GPS obtidas utilizando as funcionalidades da Google Roads API



Figura 11: Exemplo de *grid* irregular com resolução de *3 metros*

3.3 Algoritmo Proposto

O sistema proposto para fornecer a estimativa de localização foi dividido em duas etapas: a primeira destinada ao treinamento dos *grids*, também chamada de etapa *offline* e uma segunda etapa conhecida como *online*, sendo essa última, dedicada a efetuar previsões de coordenadas com base no valor proveniente de novas medições.

Para efeitos desse estudo, os dados oriundos das medições e armazenados no *reference DB* foram utilizados para treinar os tipos de *grid* apresentados na Seção anterior. Foram reservadas 100 amostras para etapa de validação do modelo de localização proposto, sendo as demais 2.656 amostras, utilizadas como referência para treinar cada uma das coordenadas presentes nos *grids*.

A Figura 12 (a) mostra em azul os pontos separados para o conjunto de teste. Essas amostras foram reservadas para validar o erro associado a estimativa de localização fornecida pelo sistema. Já a Figura 12 (b) detalha em amarelo os pontos associados as medições que compõem o conjunto de treino, os pontos em azul, correspondem aqueles do conjunto de teste mostrado na Figura 12 (a). As amostras em amarelo foram utilizadas para treinar as medições de *RSS fingerprint* associadas as coordenadas dos *grids*.

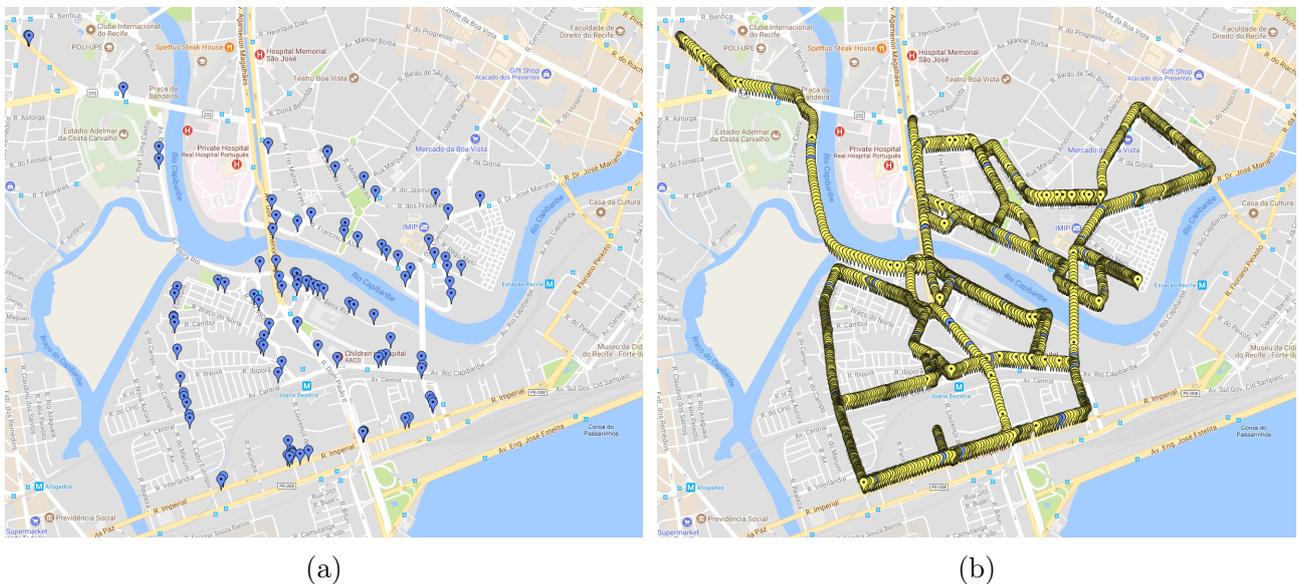


Figura 12: Conjunto de dados para treinamento e teste dos *grids*

A etapa de treinamento dos *grids* foi realizada utilizando o algoritmo WkNN (*Weighted k Nearest Neighbors*). Esse algoritmo foi selecionado devido a sua simplicidade de implementação, aliado aos trabalhos que apontam para obtenção de bons resultados quando utilizado para solução do problema aqui contextualizado [9, 19, 10].

Para minimizar a influência dos algoritmos de aprendizagem de máquina na etapa de treinamento dos *grids*, o valor das variáveis W e K foram fixados em 1.0 e 4.0 ($W = 1.0$; $K = 4.0$). Essas variáveis por sua vez, correspondem respectivamente ao peso (*weight*) e ao número de vizinhos (*k neighbors*).

A Figura 13 mostra o diagrama em blocos para o sistema de localização proposto nesse documento. É possível perceber a segmentação entre as etapas de treinamento *offline* e predição *online*. Esse sistema foi utilizado para condução dos experimentos e conseqüente recolhimento dos dados que fundamentaram a fase de análise.

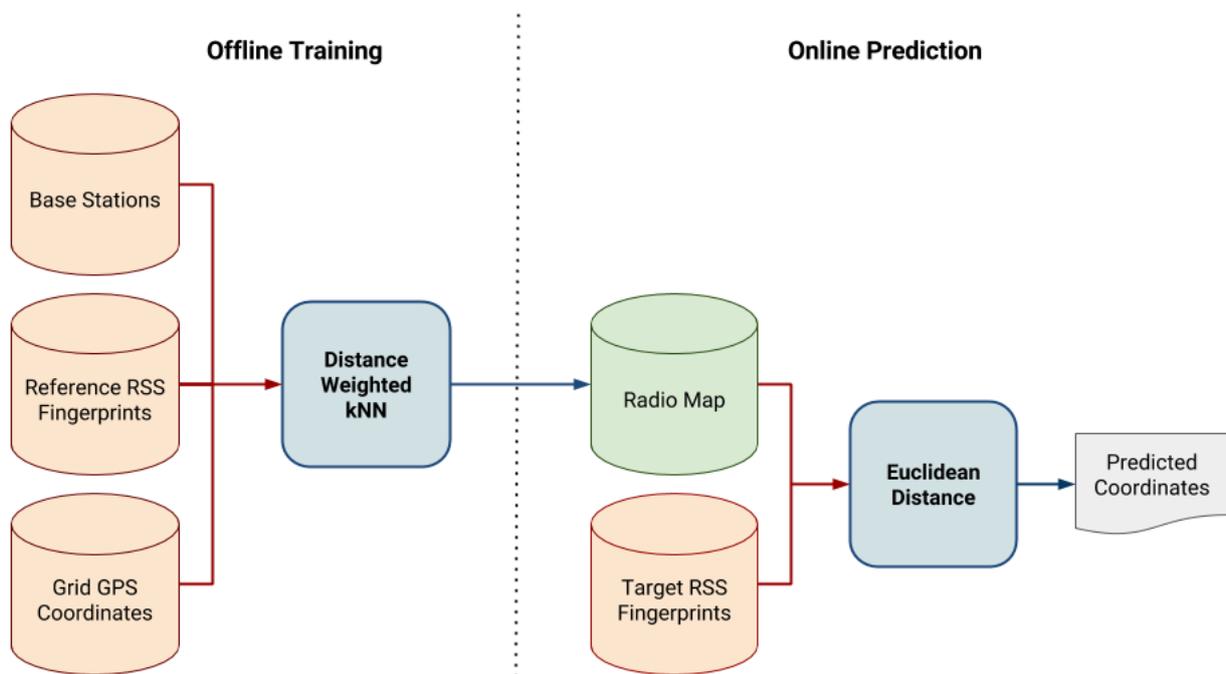


Figura 13: Diagrama em blocos para o sistema de localização

Ainda sobre a Figura 13, vale ressaltar que o *Reference RSS fingerprints* corresponde ao conjunto de amostras designadas para etapa de treinamento. Já o *Target RSS Fingerprints* corresponde as 100 amostras reservadas para validar a precisão da estimativa de localização fornecida pelo sistema. Vale lembrar que ambos conjuntos de dados foram extraídos do *reference DB* criado durante a etapa de coleta [6]. O conjunto *Base Stations* contém a localização e demais informações a respeito das estações base.

Um *Radio Map* único foi gerado para cada modelo de *grid* proposto. A base de dados *Grid GPS Coordinates* corresponde as coordenadas GPS para cada tipo de *grid*. Essa base foi substituída em cada experimento conforme o modelo de *grid* selecionado. Foram criados *grids* regulares e irregulares variando o parâmetro resolução no intervalo compreendido entre 1 e 20

metros. Como cada *Radio Map* está ligado diretamente a escolha do *grid*, esses foram sendo gerados individualmente conforme o tipo de *grid* escolhido para o experimento.

Para etapa de predição, foi utilizada a distância Euclidiana como medida numérica de similaridade entre as medições de RSS provenientes do conjunto de dados para validação (*Target RSS Fingerprints*) com aquelas presentes no mapa de rádio (*Radio Map*).

Entre os dados recolhidos durante a etapa de condução dos experimentos podemos citar: tempo de duração para o treinamento dos *gids*, tempo de duração associado a etapa predição, consumo de memória volátil em ambas as etapas, erro associado a estimativa de localização fornecida pelo sistema, entre outros. Essas informações foram agrupadas juntamente com os dados disponíveis a respeito dos *grids* para análise dos resultados obtidos.

4 Resultados

Esse Capítulo discorre a respeito dos resultados dos experimentos realizados com base no sistema de localização idealizado anteriormente. Os dados foram coletados com o propósito de tecer um comparativo entre ambos modelos de *grid*: regular e irregular.

As comparações foram efetuadas em quatro domínios distintos: *Tamanho*, *Tempo*, *Memória* e *Erro*. Essas variáveis foram analisadas com dependência no parâmetro *resolução* e no *número de células* presentes em cada modelo de *grids*.

No que concerne a análise realizada, primeiramente avaliamos os *grids* quanto ao seu tamanho, ou seja, a quantidades de células presentes em cada modelo (regular ou irregular). Em seguida relacionamos o consumo de memória volátil e o tempo de duração necessário para conclusão das etapas de treinamento (*offline*) e predição (*online*). Por último, comparamos o erro associado as estimativas de localização fornecidas em ambos os modelos.

4.1 Tamanho dos *Grids*

A palavra *tamanho* é utilizada nesse contexto para se referir a quantidade de células presentes no *grid*, quanto maior o *tamanho*, maior será a quantidade de células. Sendo assim, a primeira métrica avaliada foi o número de células em cada um dos *grids* criados durante a realização dos experimentos.

Uma característica notória a respeito dos *grids* é que seu *tamanho* é sempre o mesmo e não varia entre as etapas de treinamento e predição. Ou seja, a quantidade de células presentes em cada um dos *grids* avaliados se mantém constante no decorrer de todo o processo para obtenção de uma estimativa de localização.

A quantidade de células resultantes após o processo de criação de um *grid* é parametrizada através de uma variável chamada *resolução*. Esse parâmetro influencia diretamente o tamanho dos *grids*: quanto menor a resolução maior será a quantidade de células no *grid* construído. Nessa Seção foi realizado um comparativo a respeito do crescimento do número de células no *grid* para diferentes valores do parâmetro *resolução*.

Quando trata-se de *grids* regulares, a literatura sugere para o parâmetro *resolução* a utilização dos seguintes valores: 5m, 10m, 15m ou 20m [5]. A utilização desses valores está atrelada a premissa de que a diminuição gradativa da *resolução* acarreta em um aumento exponencial da quantidade de células no *grid*.

Para uma melhor compreensão essa característica foi expressa visualmente podendo ser observada na Figura 14. Em (a) temos um *grid* regular de 6m x 6m. Esse *grid* encontra-se dividido segundo uma *resolução* de 3m, fato esse que resulta em quatro células quadradas. Uma redução na *resolução* de 3m para 2m, como mostrado em (b), acarretaria em um aumento de 5 novas células. Uma nova redução na *resolução* de 2m para 1m levaria a um incremento de 9 vezes o número de células originais, conforme indicado em (c).

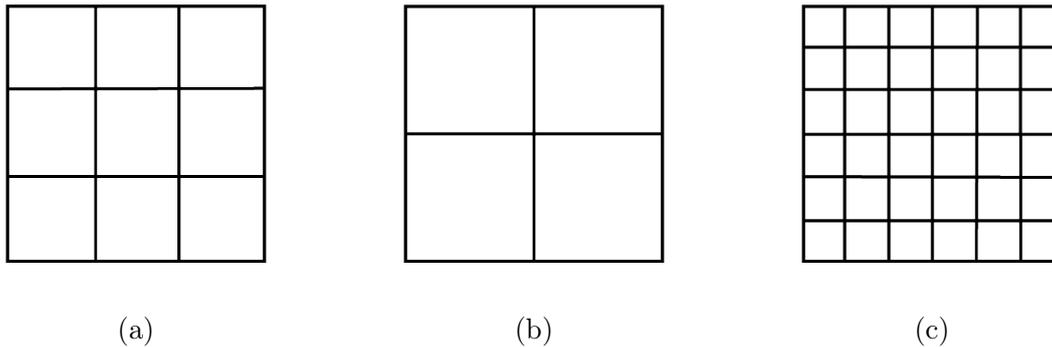


Figura 14: Influência da *resolução* na quantidade de células presentes no *grid*.

Esse aumento na quantidade de células não acarreta, ao menos a princípio, em uma melhoria significativa da precisão da estimativa de localização fornecida pelo sistema. A diminuição do erro, por si só, não seria expressiva o suficiente para justificar a adoção de valorações intermediárias para o parâmetro *resolução* tais como: 2m, 13m, 18m.

Como o estudo em questão propõe a utilização de *grids* irregulares, optamos por investigar a influência da diminuição do parâmetro *resolução* na precisão da estimativa e na quantidade de células dos *grids*, fornecendo um comparativo entre ambos os modelos. As medições relacionadas a precisão da estimativa serão tratadas posteriormente em uma Seção dedicada a essa temática.

A Tabela 2 apresenta os *grids* que foram criados para esse estudo. Estão indicados na tabela o modelo do *grid*, a resolução utilizada e a quantidade de células geradas com base no valor definido para esse parâmetro .

A Figura 15 exhibe graficamente a relação entre a quantidade de células criadas e o parâmetro *resolução*. Em vermelho encontra-se destacado o comportamento do *grid* irregular, já em azul podemos observar a curva associada ao *grid* regular. A quantidade de células em ambos os modelos de *grid* parecem descrever uma curva exponencial com relação ao parâmetro *resolução*, apesar dessa característica, sobrepondo as duas curvas percebemos que a quantidade de células criadas para os *grids* regulares é superior numericamente em todos os casos, chegando a se distanciar do *grid* regular a uma taxa cada vez maior a medida que a *resolução* diminui.

Número de Células nos *Grids*

Resolução	Regular	Irregular
1 metro	4.231.360 células	15.365 células
2 metros	1.058.224 células	8.072 células
3 metros	470.492 células	5.626 células
4 metros	264.792 células	4.419 células
5 metros	169.273 células	3.693 células
6 metros	117.750 células	3.210 células
7 metros	86.349 células	2.849 células
8 metros	66.316 células	2.619 células
9 metros	52.250 células	2.426 células
10 metros	42.300 células	2.271 células
11 metros	35.055 células	2.138 células
12 metros	29.516 células	2.032 células
13 metros	25.230 células	1.938 células
14 metros	21.574 células	1.864 células
15 metros	18.875 células	1.810 células
16 metros	16.638 células	1.756 células
17 metros	14.763 células	1.700 células
18 metros	13.104 células	1.659 células
19 metros	11.781 células	1.612 células
20 metros	10.622 células	1.575 células

Tabela 2: Quantidade de células em cada modelo de *grid* conforme a *resolução*

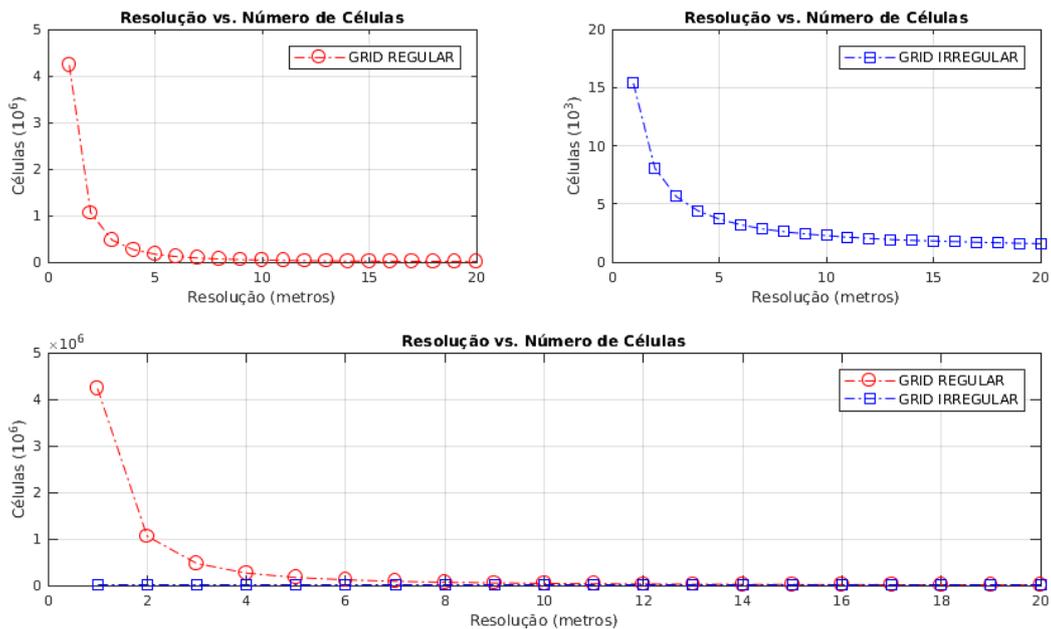


Figura 15: Relação entre a resolução dos *grids* e a quantidade de células

4.2 Tempo de Duração

Outro aspecto avaliado foi o tempo de duração relacionado as etapas treinamento dos *grids* (*offline*) e predição da localização (*online*) com base respectivamente nos conjuntos de treino e teste definidos para realização dos experimentos. Destacamos que os dados de tempo foram normalizados no intuito de fornecer informações agnósticas em relação as tecnologias ou plataforma adotada para condução desse estudo.

4.2.1 Etapa de Treinamento (*offline*)

O tempo de duração para conclusão da etapa de treinamento é função da quantidade amostras de *Reference RSS Fingerprints*. Em todas as execuções realizadas para esse experimento o tamanho do *reference DB* foi fixado em 2.656 amostras. Conforme indicado no Capítulo anterior, esse valor foi definido após a segregação da base de dados em dois conjuntos, um dedicado para etapa de treinamento, outro para realização dos testes.

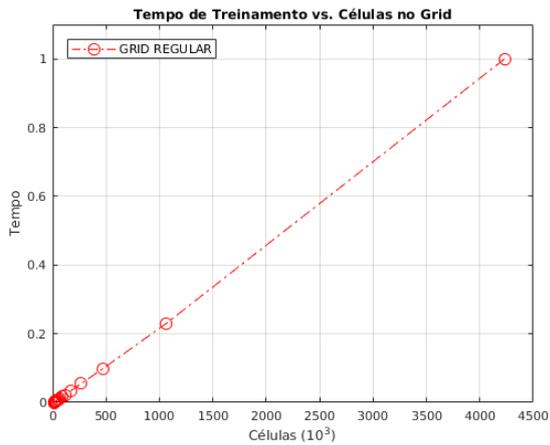
Como a quantidade de amostras consideradas na etapa de treinamento não varia entre os experimentos realizados, o fator predominante no tempo de duração dessa etapa passa a ser a quantidade de células presentes em cada modelo de *grid* analisado. Por sua vez, a quantidade de células criadas em cada *grid* é função da variável *resolução*.

Levando em consideração essas observações e no intuito de produzir um comparativo entre ambos os modelos de *grid* foram propostos alguns gráficos. O primeiro mostrado na Figura 16 revela o tempo de duração da etapa de treinamento em função da quantidade de células em cada modelo de *grid* individualmente.

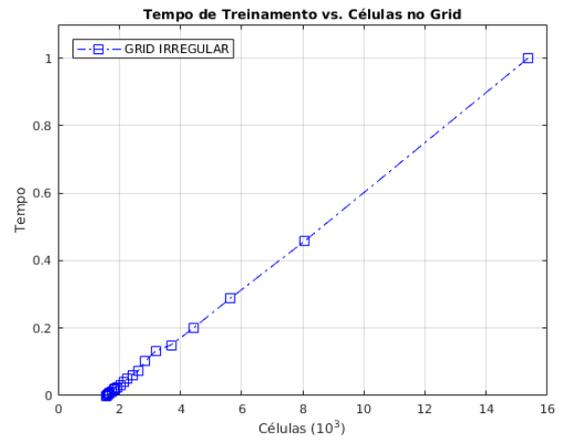
A Figura 16 (a) mostra a taxa de crescimento associada ao tempo de treinamento no modelo de *grid* regular, um procedimento análogo foi realizado para o *grid* irregular mostrado na Figura 16 (b). Em ambos os modelos a taxa de crescimento do tempo de duração parece ser linear com relação ao número de células. As equações para a taxa de crescimento do *grid* regular (1) e irregular (2) foram expressas com base nas observações realizadas.

$$T_{treino}(N_{células}) = 0.0002396 \times N_{células} + 0.002426 \quad (1)$$

$$T_{treino}(N_{células}) = 0.07118 \times N_{células} - 0.108 \quad (2)$$



(a)



(b)

Figura 16: Tempo de duração associado a etapa de treinamento dos *grids*

Como esclarecido anteriormente o parâmetro *resolução* é utilizado para criação dos *grids*, sendo o responsável pela quantidade resultante de células geradas em cada um deles. Essa variável surge como um fator comum para comparação entre ambos os modelos de *grid*.

Na Figura 17 encontram-se sobrepostos os gráficos para o tempo de duração da etapa de treinamento em função do parâmetro *resolução*. Nesse caso particular, a função resultante é decrescente já que a diminuição na resolução provoca um conseqüente aumento do número de células no *grid*. É possível visualizar o comportamento associado aos *grids* regular e irregular respectivamente em vermelho e azul.

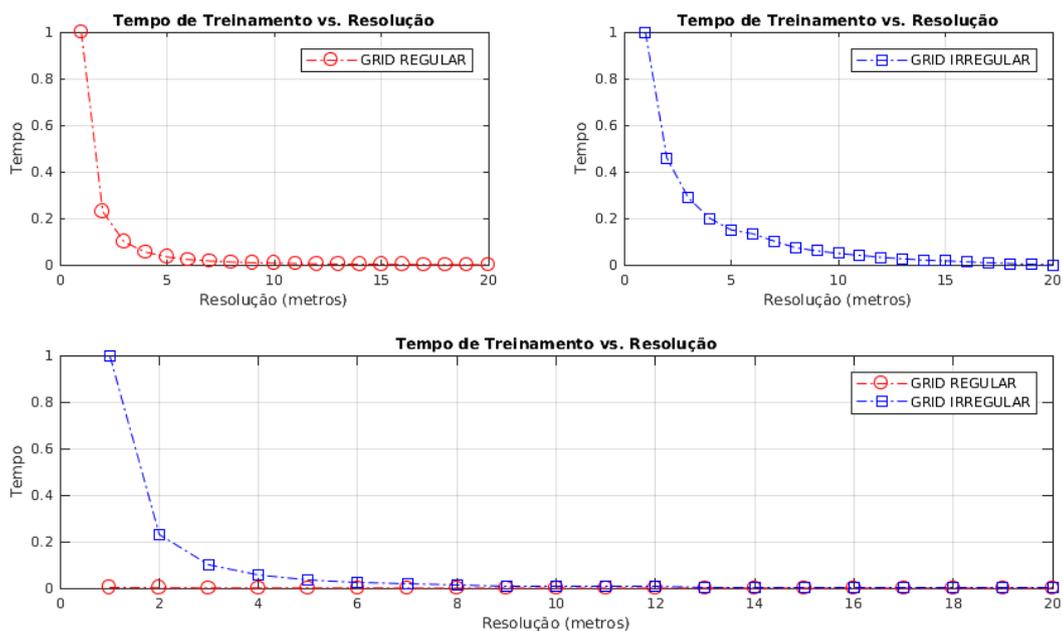


Figura 17: Relação entre o tempo de treinamento e a resolução dos *grids*

Avaliando sob a perspectiva da variável *resolução*, a curva para ambos os modelos de *grid* pode ser aproximada por uma função exponencial decrescente. As equações do tempo de duração para etapa de treinamento foram expressas com base no parâmetro *resolução* e encontram-se indicadas para o *grid* regular em (3) e para o *grid* irregular em (4).

$$T_{treino}(r) = 6.183 \times e^{-2.022 \times r} + 0.2709 \times e^{-0.4029 \times r} \quad (3)$$

$$T_{treino}(r) = 2.472 \times e^{-1.398 \times r} + 0.4895 \times e^{-0.2301 \times r} \quad (4)$$

Por último, vale ressaltar que o tempo de duração para etapa de treinamento é limitado superiormente pela complexidade do algoritmo utilizado. Para efeitos dos resultados aqui apresentados a etapa foi conduzida utilizando o algoritmo *Weighted k Nearest Neighbor* (WkNN).

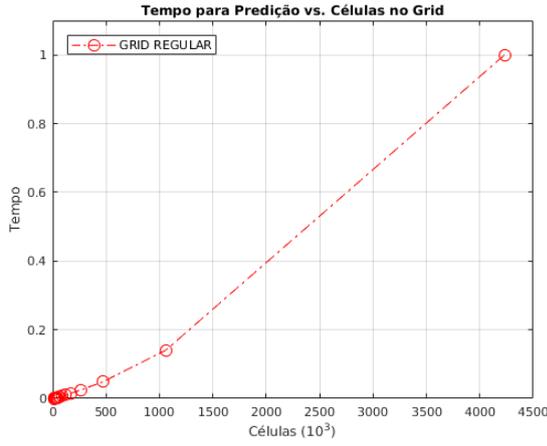
4.2.2 Etapa de Predição (*online*)

A exemplo do que acontece para a etapa de treinamento, na etapa de predição o tempo de duração é função do número de *Target RSS Fingerprints*. A exemplo do que fora estabelecido para etapa de treinamento, a quantidade de amostras desse conjunto foi fixada para todos os experimentos de modo a mitigar sua influência, restando apenas a ação do número de células presentes no *radio map*. Vale salientar ainda que o conjunto de testes fixado para utilização nos experimentos possui 100 amostras.

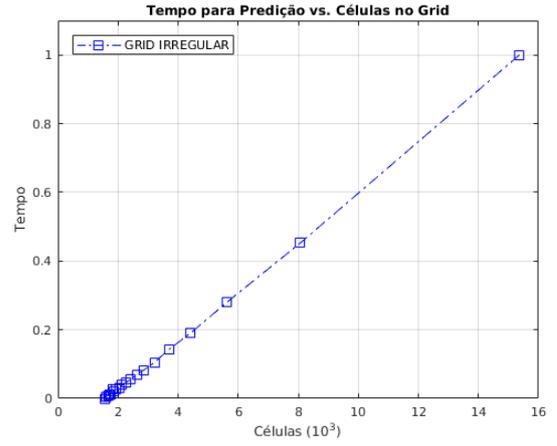
Conforme fora delineado, o valor definido para o parâmetro *resolução* influencia na quantidade de células criadas em cada modelo de *grid*. A quantidade de células presentes no *grid* corresponde ao espaço de busca analisado durante a etapa de predição. Dessa forma, assim como estabelecido para etapa de treinamento, foram efetuadas análises do tempo de duração dessa etapa para ambos os modelos de *grid* com relação a duas variáveis: *número de células* e *resolução*.

A Figura 18 revela o tempo de duração da etapa de predição em função do número de células. O tempo associado ao *grid* regular é mostrado em (a), enquanto (b) registra o tempo de duração para o *grid* irregular.

Para o gráfico mostrado na Figura 18 (a) foi efetuada uma aproximação através de uma função racional do tipo $P(x)/Q(x)$. Já para o gráfico exibido na Figura 18 (b) foi realizada uma regressão linear. Os ajustamentos encontram-se retratados na Figura 21 em (a) e (b) para



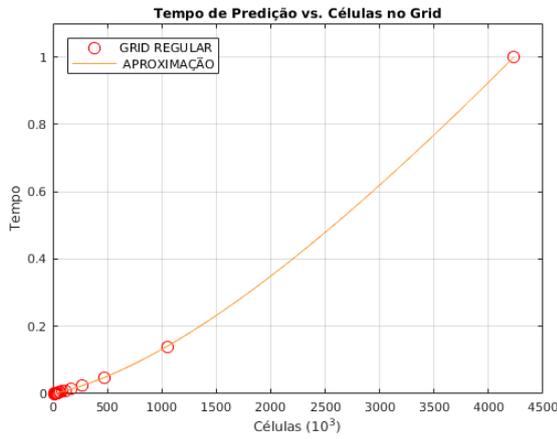
(a)



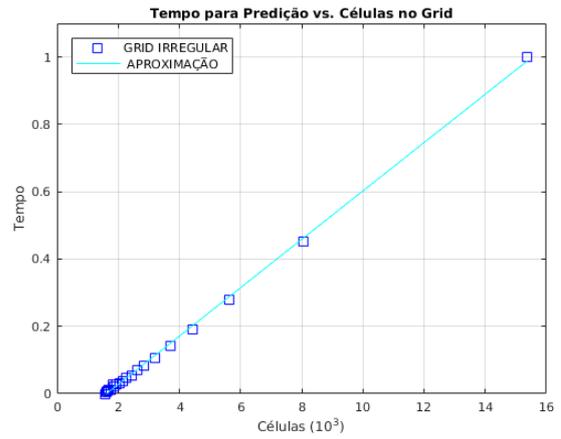
(b)

Figura 18: Tempo de duração associado a etapa de predição dos *grids*

os *grids* regulares e irregulares respectivamente. As equações associadas a essas aproximações foram expressas para o *grid* regular em (5) e para o *grid* irregular em (6).



(a)



(b)

Figura 19: Ajuste de curva para o tempo de predição com o número de células no *grid*

$$T_{teste}(N_{células}) = \frac{0.0005953 \times (N_{células})^2 + 0.0005461 \times N_{células} - 8.485}{N_{células} + 9994} \quad (5)$$

$$T_{teste}(N_{células}) = 0.07194 \times N_{células} - 0.1172 \quad (6)$$

Para facilitar a comparação com relação a variável *resolução*, os tempos de duração para ambos os modelos de *grid* foram expressos em um único gráfico mostrado na Figura 20. Através dos pontos em vermelho é possível presumir a curva associada ao *grid* regular, já os pontos em azul descrevem o comportamento do *grid* irregular.

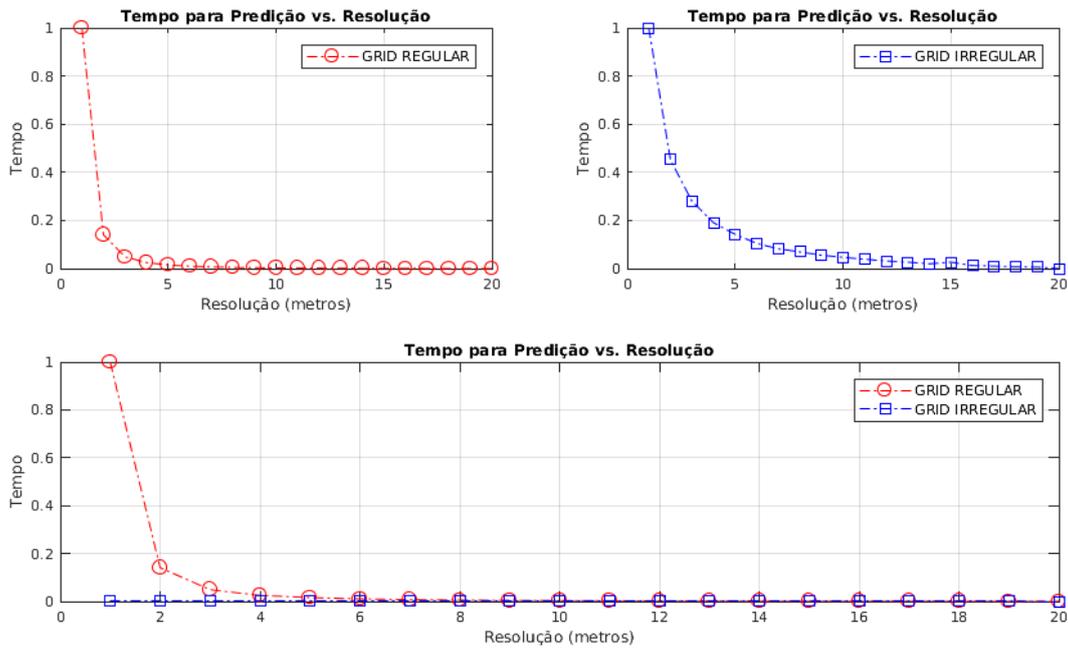


Figura 20: Relação entre o tempo de predição e a resolução dos *grids*

A Figura 21 mostra as curvas traçadas com base nos pontos para cada tipo de *grid* após efetuadas as regressões. Em vermelho visualizamos o comportamento associado ao *grid* regular (a), enquanto em azul está traçada a curva descrita pelos pontos associados ao *grid* irregular (b). As equações com base na variável independente *resolução* foram expressas em (7) e (8) para os *grids* regular e irregular respectivamente.

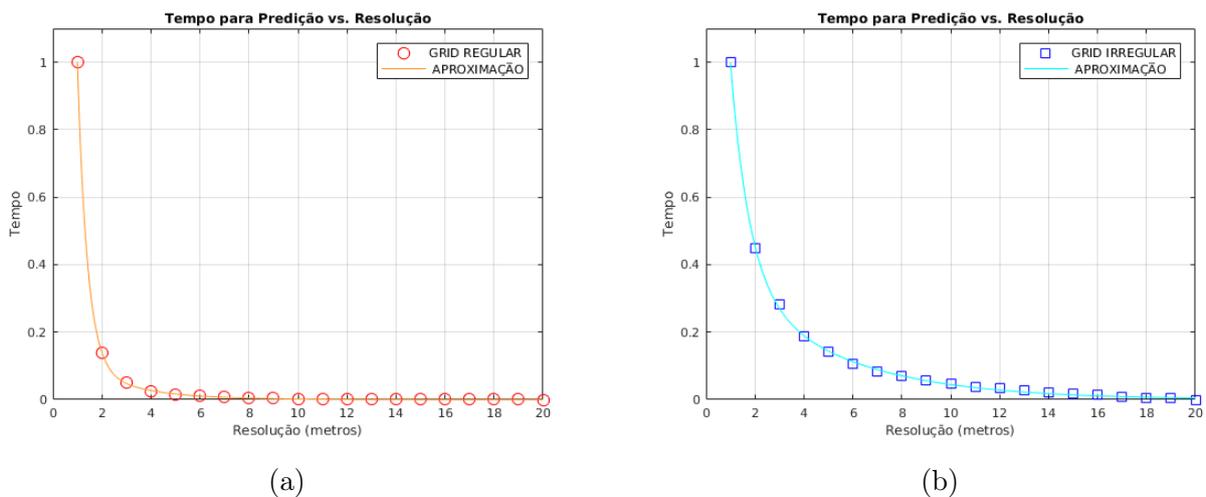


Figura 21: Ajuste de curva para o tempo de predição com o parâmetro *resolução*

$$T_{teste}(r) = 10.74 \times e^{-2.483 \times r} + 0.1631 \times e^{-0.4552 \times r} \quad (7)$$

$$T_{teste}(r) = 2.374 \times e^{-1.299 \times r} + 0.4424 \times e^{-0.2294 \times r} \quad (8)$$

Conforme mencionado anteriormente, outro fator determinante é a complexidade temporal inerente ao algoritmo escolhido. Para a etapa de predição foi utilizada a distância Euclidiana como medida de similaridade entre uma amostra de teste com aquelas presentes no *radio map*.

4.3 Consumo de Memória Volátil

As métricas a respeito do consumo de memória volátil seguem a mesma linha de análise daquela utilizada para obtenção das métricas relacionadas ao tempo. Assim como efetuado para o tempo de duração, as informações sobre o consumo de memória volátil foram obtidas para as etapas de treinamento dos *grids* e predição do posicionamento. Além disso, a normalização dos dados foi considerada no intuito de tornar a análise imparcial com relação as tecnologias e plataforma empregada na sua obtenção.

O consumo de memória volátil para cada um dos experimentos foi mensurado utilizando o valor para o *Resident Set Size* (RSS). Essa métrica não corresponde exatamente a quantidade de memória total ocupada pelo processo mas oferece uma boa aproximação. São excluídas do cálculo regiões de memória compartilhada ou paginadas para a memória secundária através de operações de *swap*.

Outra reflexão importante trazida da Seção anterior indica que o tempo de duração é função da quantidade de células presentes nos *grids*. Similarmente espera-se que o consumo de memória volátil também seja função da quantidade de células, já que quanto maior o *grid*, maior é a necessidade de carregar estruturas em memória para efeitos de atuação dos algoritmos.

Como a variável *número de células* é diretamente influenciada pelo valor do parâmetro *resolução*, esse último também foi considerado na análise realizada. Assim como efetuado na avaliação do tempo de duração, o parâmetro foi utilizado para comparação de desempenho entre os modelos de *grid* regular e irregular.

Por último, recapitulando novamente a Seção anterior, destacamos que a etapa de treinamento foi realizada utilizando 2.656 amostras de *RSS fingerprints* da base de dados de referência. Já a etapa de predição foi conduzida utilizando 100 amostras de *RSS fingerprints* oriundas do conjunto de dados separado originalmente para fase de testes.

4.3.1 Etapa de Treinamento (*offline*)

A Figura 22 mostra o consumo de memória volátil em função do número de células para ambos modelos de *grid*. Em (a) temos o comportamento descrito para os *grids* regulares, enquanto em (b) podemos verificar o padrão descrito pelos *grids* irregulares com o aumento do número de células.

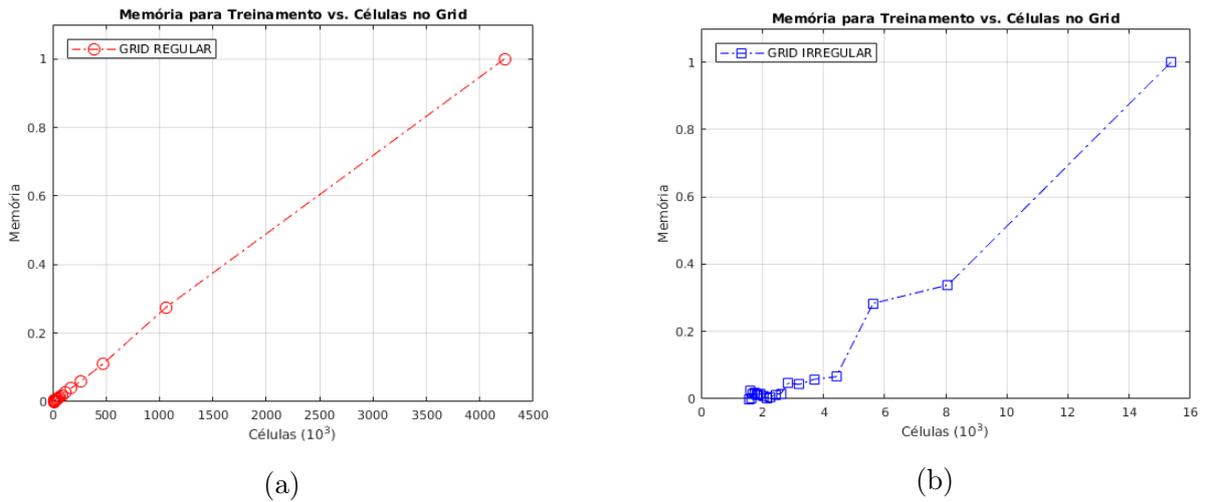


Figura 22: Consumo de memória volátil com o aumento do número de células

A Figura 23 indica o ajustamento de curva realizado para ambos os modelos de *grid*. Em ambos os casos a função de crescimento para os *grids* pode ser aproximada através de uma reta. Em (a) podemos visualizar a regressão para o *grid* regular, já a imagem (b) mostra a reta traçada para o *grid* irregular.

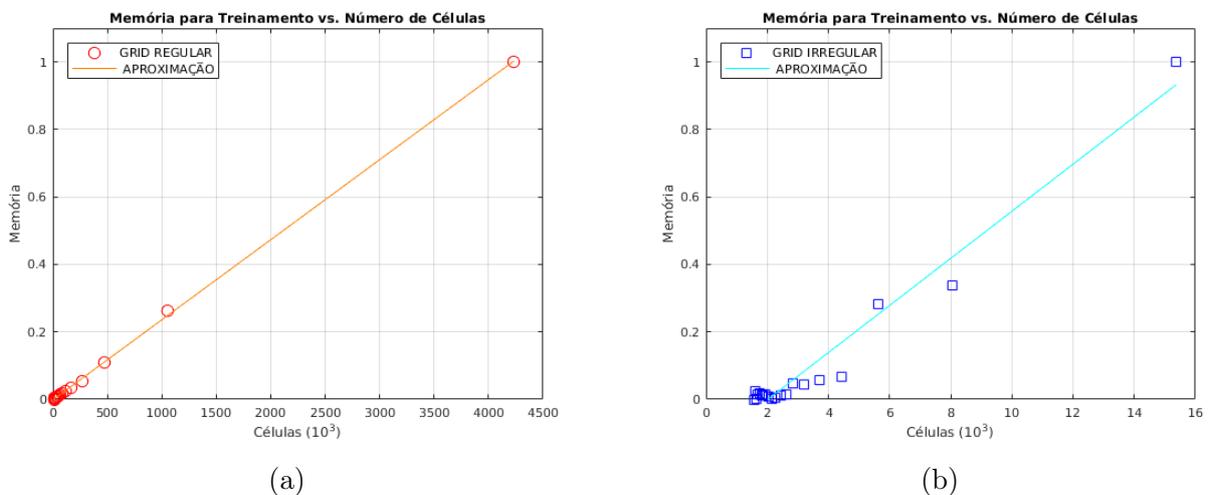


Figura 23: Ajuste de curva para o consumo de memória volátil com o número de células

Foram expressas as equações para as funções cujos contornos encontram-se desenhados na Figura 23 (a) e (b). As equações (9) e (10) mostram as relações obtidas respectivamente para os *grids* regular e irregular.

$$M_{treino}(N_{celulas}) = 0.0002373 \times N_{celulas} - 0.001513 \quad (9)$$

$$M_{treino}(N_{celulas}) = 0.06932 \times N_{celulas} - 0.04473 \quad (10)$$

Uma análise similar foi realizada com base no parâmetro *resolução*. A Figura 24 ilustra para fins de comparação o comportamento de ambos os *grids* com relação a essa variável. Em vermelho observamos o consumo de memória para o *grid* regular, já os pontos em azul descrevem o desempenho dos *grids* irregulares quanto a grandeza aqui analisada.

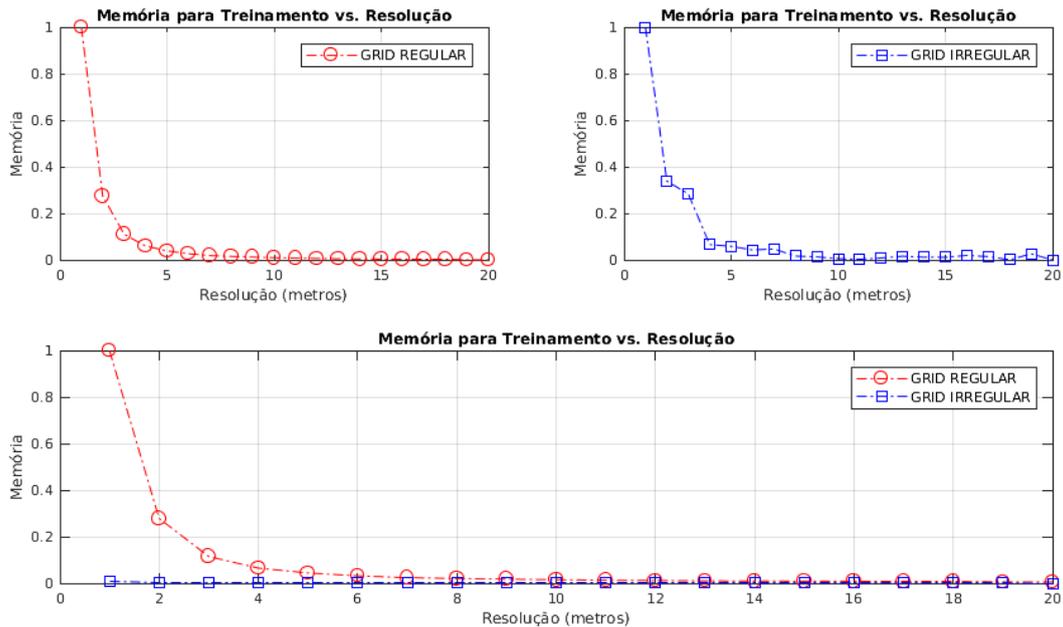


Figura 24: Relação entre a memória para treinamento e a resolução dos *grids*

Conforme indicado na Figura 25, a curva descrita para esse caso segue um padrão semelhante aos obtidos anteriormente. Em ambos os modelos de *grid* podemos efetuar uma aproximação utilizando uma função exponencial. Em (a) podemos visualizar o ajuste realizado para o *grid* regular, a imagem (b) mostra a curva descrita pelo *grid* irregular.

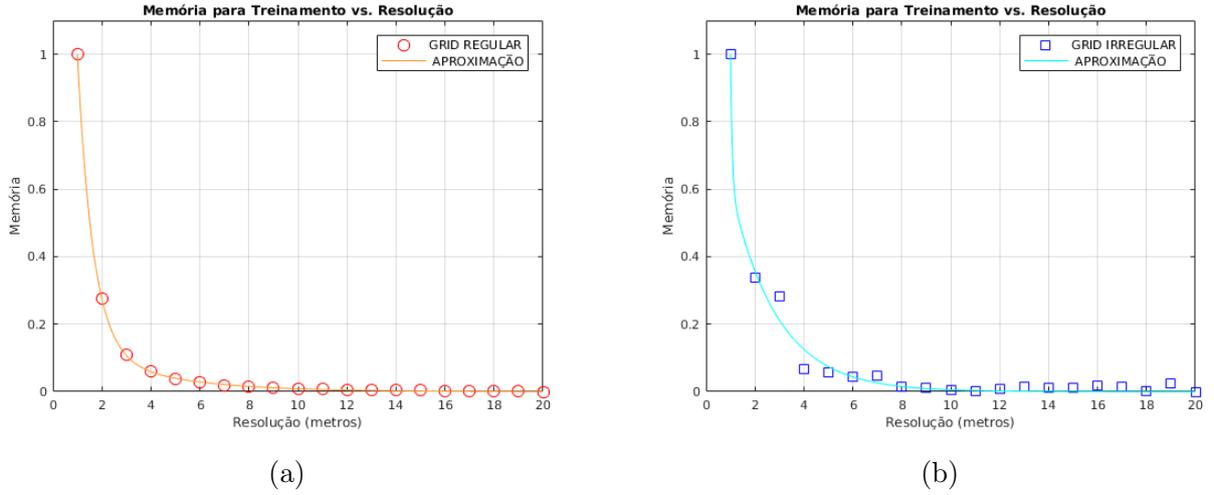


Figura 25: Ajuste de curva para o consumo de memória volátil com a resolução do *grid*

As equações que descrevem o comportamento apresentado na Figura 25 foram expressas em (11) e (12). A expressão em (11) revela a função encontrada para o *grid* regular, enquanto (12) retrata a equação obtida para o *grid* irregular.

$$M_{treino}(r) = 4.427 \times e^{-1.631 \times r} + 0.1839 \times e^{-0.3243 \times r} \quad (11)$$

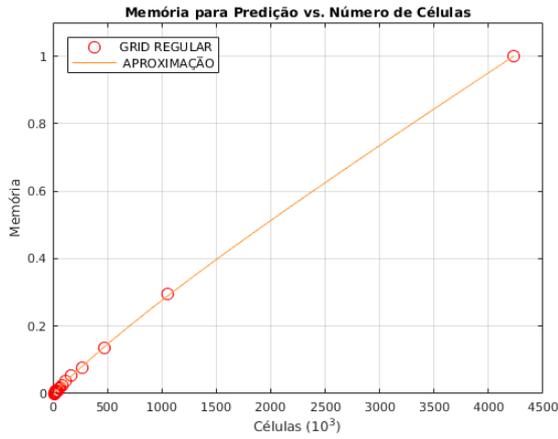
$$M_{treino}(r) = 1.944 \times e^{-1.041 \times r} + 0.3556 \times e^{-0.1054 \times r} \quad (12)$$

4.3.2 Etapa de Predição (*online*)

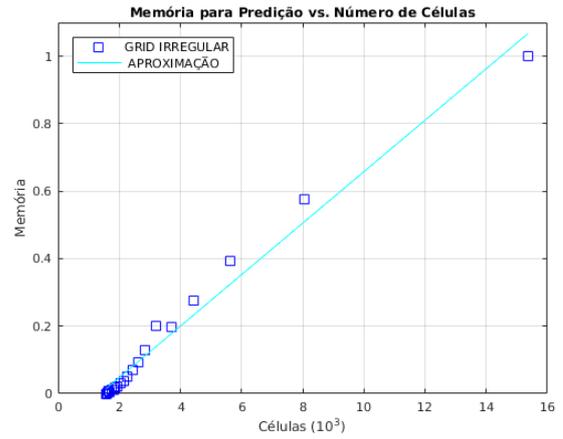
As métricas para o consumo de memória volátil na etapa de predição seguem o mesmo princípio adotado na etapa de treinamento. Apesar do comportamento semelhante, as relações que distinguem respectivamente os *grids* regulares dos irregulares ficam um pouco mais evidentes nessa etapa. As distribuições associadas ao número de células no *grid* representam de forma mais fiel o contorno das curvas obtidas, principalmente para o *grid* irregular.

A Figura 26 mostra o consumo de memória volátil com relação ao número de células em cada um dos modelos de *grid* para a etapa de predição. São mostrados também os ajustes de curva efetuados para ambos os casos. O *grid* regular é considerado em (a), enquanto (b) retrata o comportamento para o *grid* irregular.

As equações obtidas para as retas apresentadas na Figura 26 (a) e (b) foram expressas e encontram-se descritas em (13) e (14). Em (13) temos a reta para o modelos de *grid* regular, enquanto em (13) encontramos a expressão para o *grid* irregular.



(a)



(b)

Figura 26: Ajuste de curva para o consumo de memória volátil com o número de células

$$M_{teste}(N_{células}) = 0.0002381 \times N_{células} + 0.00446 \quad (13)$$

$$M_{teste}(N_{células}) = 0.076 \times N_{células} - 0.1074 \quad (14)$$

O comportamento para o consumo de memória associado a variação do parâmetro *resolução* pode ser visualizado na Figura 27. Em vermelho está representado o *grid* regular, enquanto os pontos em azul condizem com o padrão relacionado ao *grid* irregular.

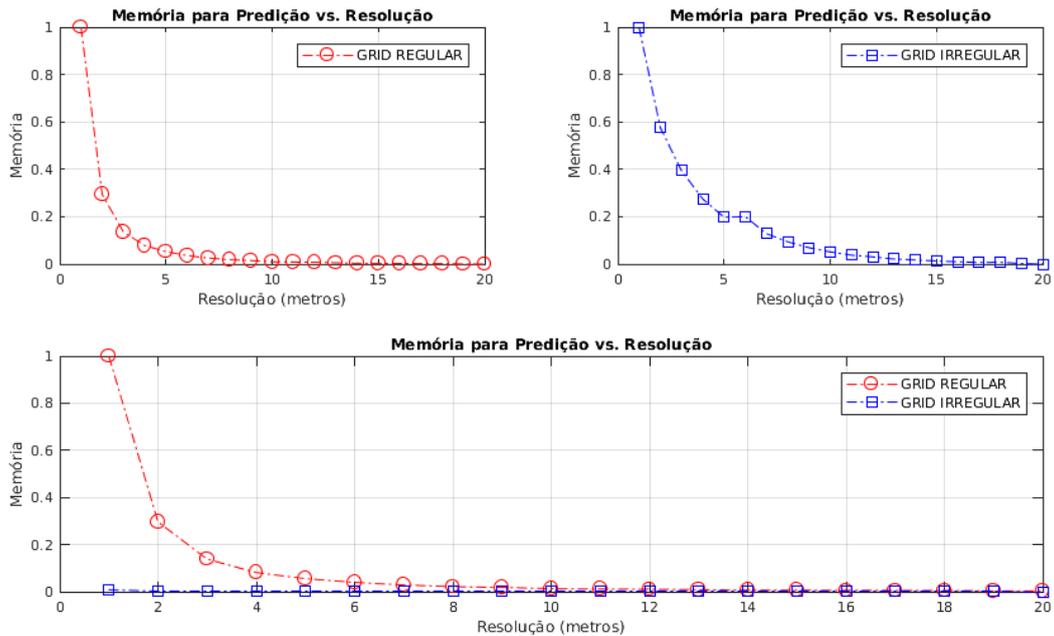


Figura 27: Relação entre a memória para teste e a resolução dos *grids*

Assim como efetuado para o número de células o ajustamento das curvas foi realizado para ambos os modelos de *grid*. As equações (15) e (16) expressam as funções obtidas respectivamente para os *grids* regular e irregular.

$$M_{teste}(r) = 10.74 \times e^{-2.483 \times r} + 0.1631 \times e^{-0.4552 \times r} \quad (15)$$

$$M_{teste}(r) = 2.374 \times e^{-1.299 \times r} + 0.4424 \times e^{-0.2294 \times r} \quad (16)$$

4.4 Erro na Estimativa

A principal métrica para avaliação de um sistema de posicionamento é a precisão na estimativa de localização. Sendo assim, as medições obtidas para o erro máximo, médio e mínimo são imprescindíveis para efeitos de comparação qualitativa entre os sistemas disponíveis.

Como delineado em Seção anterior, com a finalidade de ressaltar a importância da escolha da geometria do *grid*, os algoritmos utilizados nas fases de treinamento e predição foram preestabelecidos. O objetivo é minimizar a influência desses algoritmos fornecendo um patamar fundamental, necessário para comparação entre os dois modelos de *grid*.

Assim como efetuado para as outras métricas observadas, o valor para o erro na estimativa de localização foi obtido com relação ao parâmetro *resolução* dos *grids*. Conforme indicado na Seção 4.1, essa variável foi escolhida devido a sua popularidade na literatura.

Vale destacar que o valor para o erro foi obtido após a etapa de predição, ou seja, levamos em consideração a estimativa de localização fornecida pelo sistema para as 100 amostras utilizadas nessa etapa. Para fins da análise aqui proferida, os erros máximo e mínimo implicam respectivamente no maior e menor valor em metros obtidos através da estimativa realizada. Já o valor médio foi calculado a partir da média aritmética do erro associado a estimativa de localização produzido para as 100 amostras. A Tabela 3 consolida os resultados obtidos através dos experimentos realizados.

<i>Grid</i>			Erro	
Resolução	Geometria	Células	Médio	Desvio Padrão
1 m	<i>regular</i>	4.231.360	120,05 m	223,32 m
	<i>irregular</i>	15.365	94,03 m	223,73 m

<i>Grid</i>			Erro	
Resolução	Geometria	Células	Médio	Desvio Padrão
2 m	<i>regular</i>	1.058.224	115,47 m	224,11 m
	<i>irregular</i>	8.072	93,76 m	223,88 m
3 m	<i>regular</i>	470.492	123,40 m	222,17 m
	<i>irregular</i>	5.626	94,53 m	223,82 m
4 m	<i>regular</i>	264.792	120,77 m	221,12 m
	<i>irregular</i>	4.419	96,10 m	224,47 m
5 m	<i>regular</i>	169.273	122,62 m	226,04 m
	<i>irregular</i>	3.693	99,58 m	224,74 m
6 m	<i>regular</i>	117.750	125,83 m	224,71 m
	<i>irregular</i>	3.210	99,15 m	224,11 m
7 m	<i>regular</i>	86.349	128,83 m	223,69 m
	<i>irregular</i>	2.849	101,02 m	224,73 m
8 m	<i>regular</i>	66.316	121,65 m	225,49 m
	<i>irregular</i>	2.619	98,31 m	225,29 m
9 m	<i>regular</i>	52.250	130,35 m	226,07 m
	<i>irregular</i>	2.426	96,69 m	223,86 m
10 m	<i>regular</i>	42.300	123,70 m	224,00 m
	<i>irregular</i>	2.271	97,02 m	223,89 m
11 m	<i>regular</i>	35.055	117,56 m	220,56 m
	<i>irregular</i>	2.138	101,56 m	223,44 m
12 m	<i>regular</i>	29.516	120,54 m	221,08 m
	<i>irregular</i>	2.032	104,57 m	226,20 m
13 m	<i>regular</i>	25.230	129,47 m	226,85 m
	<i>irregular</i>	1.938	109,69 m	227,02 m
14 m	<i>regular</i>	21.574	125,01 m	222,61 m
	<i>irregular</i>	1.864	100,36 m	224,41 m

<i>Grid</i>			Erro	
Resolução	Geometria	Células	Médio	Desvio Padrão
15 m	<i>regular</i>	18.875	132,05 m	226,11 m
	<i>irregular</i>	1.810	101,16 m	224,33 m
16 m	<i>regular</i>	16.638	128,43 m	225,39 m
	<i>irregular</i>	1.756	100,85 m	224,39 m
17 m	<i>regular</i>	14.763	145,76 m	231,35 m
	<i>irregular</i>	1.700	102,02 m	223,76 m
18 m	<i>regular</i>	13.104	130,93 m	223,59 m
	<i>irregular</i>	1.659	101,98 m	223,09 m
19 m	<i>regular</i>	11.781	134,35 m	226,44 m
	<i>irregular</i>	1.612	102,02 m	222,77 m
20 m	<i>regular</i>	10.622	131,88 m	223,30 m
	<i>irregular</i>	1.575	102,53 m	222,49 m

Tabela 3: Erro associado a estimativa de localização para os modelos de *grid*

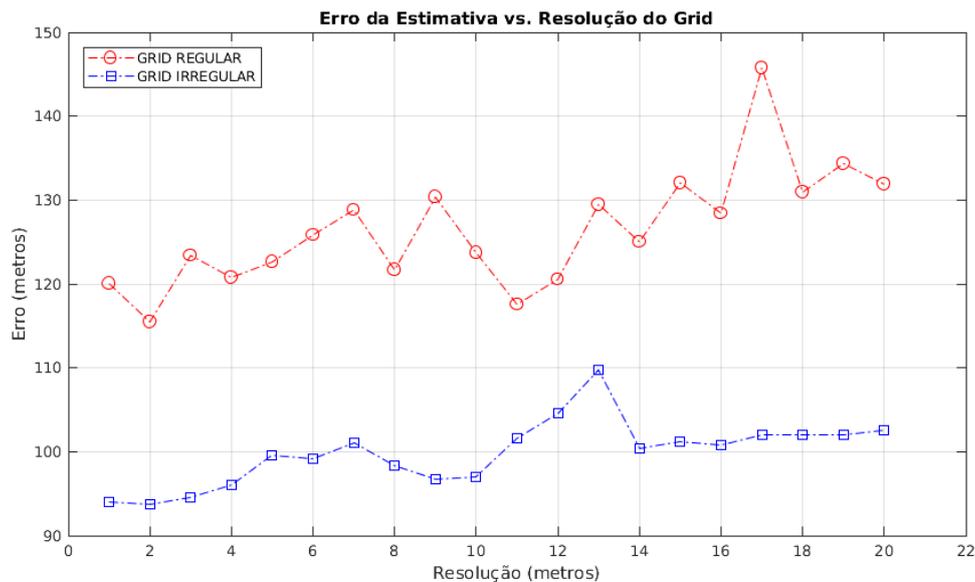


Figura 28: Comparativo entre os modelos de *grid* quanto a precisão da estimativa

Conforme visualizado na Tabela 3, o valor médio para o erro foi isolado para cada um dos experimentos realizados e expresso graficamente na Figura 28. Os pontos em vermelho estão

associados aos *grids* de geometria regular, já os pontos em azul fazem referência aos *grids* cuja geometria é irregular. No gráfico proposto, o parâmetro *resolução* pode ser utilizado para estabelecer um critério de comparação entre ambos os modelos de *grid*, por essa razão, os pontos para os dois modelos de *grid* encontram-se sobrepostos no gráfico.

Através da Figura 28 é possível notar que, em todos os experimentos realizados, os *grids* de geometria irregular possuem uma precisão superior aqueles de geometria regular. Vale destacar ainda que esse ganho na precisão é obtido com uma diminuição na quantidade de células do *grid*, fato esse que acarreta em melhorias de performance quanto ao tempo de duração e ao consumo de memória volátil, tanto na etapa de treinamento dos *grids* (fase *offline*), quanto na etapa para predição da localização (fase *online*).

Apenas exemplificando, ao fixarmos a resolução dos *grids* em 1 metro, aquele de geometria irregular apresenta uma quantidade de células cerca de 275 vezes menor que o de geometria regular, passando de 4.231.360 células do *grid* regular para 15.365 no *grid* irregular. Além dessa característica, o erro na estimativa de posicionamento diminui aproximadamente 26 metros, passando de cerca de 120 metros no *grid* regular para 94 metros no *grid* irregular.

Outro aspecto observado através da Figura 28 está relacionado a tendência de diminuição do erro médio associado a estimativa de localização a medida que o parâmetro *resolução* se torna cada vez menor. Essa tendência pode ser verificada ao efetuar uma regressão linear com base nos pontos do gráfico para ambos os modelos de *grid* mostrados originalmente na imagem. A Figura 29 mostra o resultado obtido através dessa operação.

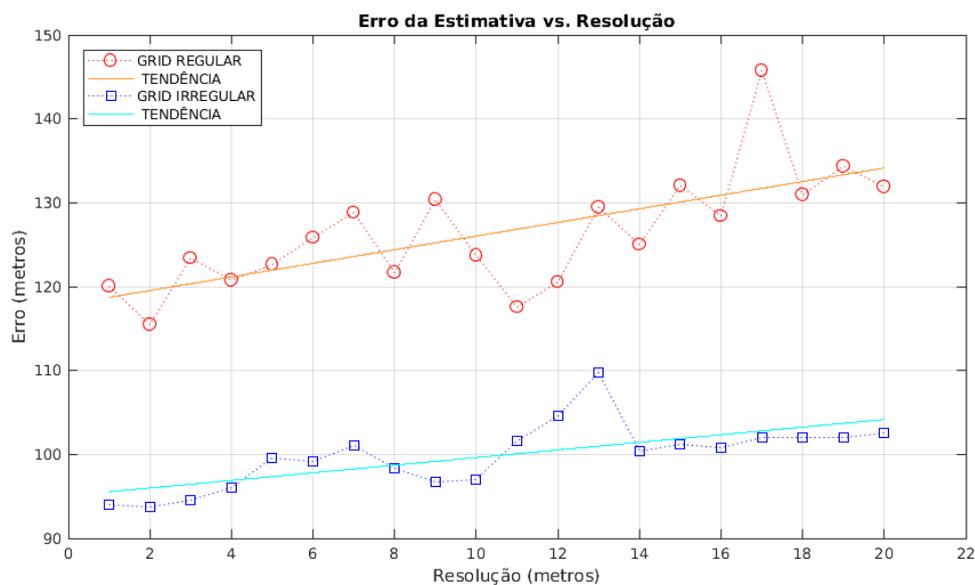


Figura 29: Tendência de diminuição do erro com relação a *resolução*

As equações de tendência obtidas para a Figura 28 encontram-se indicadas a seguir. A equação (17) descreve a tendência associada do erro para o *grid* regular, enquanto a equação (17) indica a tendência de crescimento do erro para o *grid* de geometria irregular.

$$E(r) = 0.8124 \times r + 117.9 \quad (17)$$

$$E(r) = 0.4535 \times r + 95.09 \quad (18)$$

Ainda sobre os *grids* regulares, a utilização de valores muito baixos para *resolução* do *grid*, abaixo de 5 metros por exemplo, torna-se um impeditivo devido a explosão exponencial na quantidade de células resultantes.

Apesar do aumento do número de células com a diminuição da *resolução* estar condicionada a um crescimento exponencial, conforme Seção 4.1, esse aumento é rigorosamente mais acentuado para o *grid* de geometria regular. A variação na *resolução* de 20 metros para 1 metro nesse modelo de *grid* leva a um aumento de quase 400 vezes no número de células. Já para o *grid* de geometria irregular, a mesma variação na *resolução* proporciona um aumento em cerca de 10 vezes no número de células. Essa característica permite utilizar esse modelo de *grid* com resoluções menores, sem onerar de forma significativa o desempenho do sistema de localização.

Por último, vale lembrar que o objetivo dessa análise é simplesmente fornecer um comparativo entre os dois modelos de *grid*. Um valor absoluto menor para o erro associado a estimativa de posicionamento poderia ser obtido, em ambos os casos, utilizando algoritmos como WkNN (*Weighted k Nearest Neighbors*), SVM (*Support Vector Machine*) ou até mesmo ANN (*Artificial Neural Network*) na etapa de predição [1, 2]. Por tratar-se de um estudo comparativo optamos apenas por utilizar a distância Euclidiana para aferir o valor médio do erro.

5 Conclusão

A hipótese levantada pôde ser confirmada: em cenários urbanos densamente ocupados por edificações de morfologia e distribuição arbitrárias, o problema de localização *outdoor* pode ser reduzido para um problema de identificar em qual região da malha viária urbana uma dada estação móvel encontra-se.

Essa redução no espaço de busca do problema trouxe resultados significativos na acurácia do modelo de localização proposto nesse estudo, fornecendo ganhos com relação ao tempo de duração e ao consumo de memória volátil nas etapas de treinamento e predição, além de benefícios quanto a melhoria na precisão da estimativa.

Com base nessa afirmação um questionamento surge quase que imediatamente: caso a estação móvel não esteja necessariamente na malha viária ou em seus arredores, o que aconteceria com a estimativa de localização para regiões do perímetro urbano livres de edificações? Essas regiões poderiam ser caracterizadas, por exemplo, por parques e praças dentro da cidade ou até mesmo pelo topo de algumas edificações.

Apesar desse aparente impacto no resultado das estimativas de localização, suposições otimistas baseiam-se no fato de que essas regiões são minoria nos grandes centros urbanos. Além disso, esses espaços ao ar livre são notadamente permeados por ciclovias ou travessias de pedestre, trechos esses, passíveis de serem mapeados sem gerar um aumento significativo na quantidade de pontos em um *grid* de geometria irregular.

Vale ressaltar ainda que trechos como ciclovias e passarelas são substancialmente mais relevantes para o problema de localização, ou seja, existe uma tendência maior de que estações móveis sejam encontradas sob essas vias do que nas demais regiões de uma áreas não edificada.

Outra evidência que corrobora para essa avaliação mais otimista está relacionada ao fato que, nos grandes centros urbanos, a proporção em quilômetros quadrados de malha viária é superior aquela dedicada a parques e praças, característica essa que poderia minimizar o impacto dessas regiões na precisão absoluta do sistema, já que um aumento inerente ao erro para esses ambientes é, a princípio, inevitável.

Por fim, devido ao fato das áreas urbanas serem cortadas por ruas e avenidas, é razoável afirmar que existe uma grande probabilidade de que uma dada estação móvel esteja se deslocando através dessas vias ou localizada em suas proximidades. Essa característica aponta favoravelmente para a adoção do modelo de *grid* proposto nesse estudo.

Referências

- [1] Quoc Duy Vo and Pradipta De. A survey of fingerprint-based outdoor localization. In *IEEE Communications Surveys & Tutorials*, volume 18, pages 491–506. IEEE, 2016.
- [2] Hui Liu, Houshang Darabi, Pat Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. In *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, volume 37, pages 1067–1080. IEEE, 2007.
- [3] Guolin Sun, Jie Chen, Wei Guo, and K.J. Ray Liu. Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs. In *IEEE Signal Processing Magazine*, volume 22, pages 12–23. IEEE, 2005.
- [4] M.P. Wylie and J. Holtzman. The non-line of sight problem in mobile location estimation. In *Proceedings of ICUPC - 5th International Conference on Universal Personal Communications*. IEEE, 1996.
- [5] H. Laitinen, J. Lahteenmaki, and T. Nordstrom. Database correlation method for gsm location. In *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No.01CH37202)*. IEEE, 2001.
- [6] Robson D.A. Timoteo, Lizandro N. Silva, Daniel C. Cunha, and George D.C. Cavalcanti. An approach using support vector regression for mobile location in cellular networks. In *Computer Networks*, volume 95, pages 51–61, 2016.
- [7] Jae-Hoon Kim, Kyoung Sik Min, and Woon-Young Yeo. A design of irregular grid map for large-scale wi-fi lan fingerprint positioning systems. In *The Scientific World Journal*, volume 2014, pages 1–13, 2014.
- [8] Yanan Han, Huan Ma, and Lijun Zhang. An efficient rf fingerprint positioning algorithm based on uneven grid layout. In *6th International Conference on Wireless, Mobile and Multi-Media (ICWMMN 2015)*, 2014.
- [9] Rafael Saraiva Campos and Lisandro Lovisolo. Location methods for legacy gsm handsets using coverage prediction. In *2008 IEEE 9th Workshop on Signal Processing Advances in Wireless Communications*. IEEE, 2008.
- [10] Marina Vuckovic, Ivan Petrovic, Deni Vidovic, Zarko Kostovic, Szilveszter Pletl, and Dragan Kukolj. Space grid resolution impact on accuracy of the indoor localization fingerprinting. In *2011 19th Telecommunications Forum (TELFOR) Proceedings of Papers*. IEEE, 2011.

- [11] In *IEEE Standard Definitions of Terms for Antennas*. IEEE.
- [12] Yuvraj Singh. Comparison of okumura, hata and cost-231 models on the basis of path loss and signal strength. In *International Journal of Computer Applications*, volume 59, pages 37–41. IEEE, 2012.
- [13] Digital mobile radio towards future generation systems. In *European Communities, EUR 18957*. EUR, 1999.
- [14] The analysis of the coexistence of fwa cells in the 3.4 - 3.8 ghz band. In *Electronic Communication Committee (ECC) within the European Conference of Postal and Telecommunications Administration (CEPT)*, number 33. ECC, 2003.
- [15] V. Erceg K. V. S. Hari et al. Channel models for fixed wireless applications. In *IEEE 802.16 Broadband Wireless Access Working Group*. IEEE, 2001.
- [16] In *ETSI TS 100911 v6.2.0 (1998 - 07) — Digital Cellular Telecommunications System (Phase 2 +); Radio Subsystem Link Control (GSM 05.08 version 6.2.0 Release 1997)*. European Telecommunications Standard Institute, 1998.
- [17] In *IEEE Standard for Information Technology — Telecommunications and Information Exchange between Systems — Local and Metropolitan Area Networks — Specific Requirements; Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. IEEE Computer Society, 2007.
- [18] In *Introduction to the Google Maps Roads API*. Google Developers.
- [19] Rafael Saraiva Campos and Lisandro Lovisolo. A fast database correlation algorithm for localization of wireless network mobile nodes using coverage prediction and round trip delay. In *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*. IEEE, 2009.