



Ministério da
Ciência e Tecnologia



sid.inpe.br/mtc-m19/2010/11.11.13.41-RPQ

IDENTIFICAÇÃO DE REGIÕES GEOGRÁFICAS DE ORIGEM E DE DESTINO DE TRÁFEGOS AÉREOS DESCONHECIDOS DO SISTEMA DE CONTROLE DO TRÁFEGO AÉREO BRASILEIRO

André Luís Maia Baruffaldi

Relatório de Pesquisa da Disciplina Mineração de Dados, orientado pelo Prof. Dr.
Rafael dos Santos, aprovada em 20 de outubro de 2010.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/38JDMDE>>

INPE
São José dos Campos
2010

PUBLICADO POR :

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):

Presidente:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Membros:

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr^a Regina Célia dos Santos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr. Horácio Hideki Yanasse - Centro de Tecnologias Especiais (CTE)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Deicy Farabello - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Vivéca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)



Ministério da
Ciência e Tecnologia



sid.inpe.br/mtc-m19/2010/11.11.13.41-RPQ

IDENTIFICAÇÃO DE REGIÕES GEOGRÁFICAS DE ORIGEM E DE DESTINO DE TRÁFEGOS AÉREOS DESCONHECIDOS DO SISTEMA DE CONTROLE DO TRÁFEGO AÉREO BRASILEIRO

André Luís Maia Baruffaldi

Relatório de Pesquisa da Disciplina Mineração de Dados, orientado pelo Prof. Dr.
Rafael dos Santos, aprovada em 20 de outubro de 2010.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/38JDMDE>>

INPE
São José dos Campos
2010

RESUMO

Identificar regiões, no território brasileiro, onde ocorram repetidas detecções ou perdas de detecções radar de aeronaves que não foram reconhecidas pelo sistema de controle do tráfego aéreo. Com base nos dados fornecidos pelo Comando da Aeronáutica, realizar um processo de estimação de posições futuras após a última identificação radar, considerando que a aeronave saiu da cobertura radar, devido a pouso. Processo semelhante é feito antes da primeira identificação, supondo que a aeronave possa ter decolado de uma pista próxima. As regiões com maior número de primeiras ou últimas detecções ou posições estimadas são as candidatas a uma análise mais detalhada à procura de pista de pouso. Tendo em vista que as detecções e as posições estimadas não são origem nem destino e que o meio de transporte utilizado é o aéreo, com grandeza de velocidade significativa, concluiu-se que o processo deve trabalhar com regiões geográficas ao invés de pontos. O objetivo é restringir áreas de interesse para que aplicativos de identificação de padrões localizem automaticamente locais que possam ser utilizados para decolagem e pouso de aeronaves, complementando os locais já catalogados. O método demonstra ser eficiente quando, após identificar regiões densas de pontos, realiza-se uma procura manual em imagens de satélite e identifica-se pistas de pouso. O método não é conclusivo, pois não é possível afirmar que a aeronave decolou ou pousou em determinado aeródromo, mas somente que existe a possibilidade de tê-lo realizado. A existência de cluster não significa absolutamente que existe pista de pouso e de decolagem, mas que há perda de detecção na região.

ABSTRACT

Identify regions in Brazil, where occur repeated detections or losses detections of radar aircraft, that were not recognized by the system of air traffic control. Based on data provided by the Air Force Command, conduct a process of estimating future positions after the last radar identification, considering that the aircraft went out of radar coverage, due to landing. A similar process is done before the first identification, assuming that the aircraft can take off from a runway nearby. The regions with the highest number of first or last detections or estimated positions are candidates for a more detailed search of the airfield. Considering that the detections and the estimated positions are not origin or destination and the transport used is by air, with significant speed magnitude, so the process should to work with geographic regions instead of points. The goal is to restrict areas of interest for applications of pattern recognition to automatically locate sites that may be used for takeoff and landing aircraft, complementing the previously cataloged. The method proves to be effective when, after identifying the dense regions of points, carried out a manual search on satellite images and identifies runways. The method is not conclusive, because you cannot say that the aircraft took off or landed at a certain airfield, but only that it is possible to have it done. The existence of cluster does not mean that there is absolutely runway and takeoff, but that a loss of detection in the region.

LISTA DE FIGURAS

	Pág.
Figura 1 – Sequência do processo de mineração de dados	2
Figura 2 – Convergência de 4 rotas para Maringá	4
Figura 3 – Rota simulada entre Curitiba e São Paulo	7
Figura 4 – Tabela de agrupamentos, contendo posição inicial, posição final e extrapolações	8
Figura 5 – Quadrículas de 60NM e de 30NM	9
Figura 6 – Visualização de quadrícula e agrupamento de rotas	10
Figura 7 – aproximação da Fig 6, onde são observadas 2 pistas de pouso.	10
Figura 8 – Superposição de rotas e quadrículas	10
Figura 9 – Procura manual por pista de pouso na Figura 8	11
Figura 10 – Pista de pouso localizada em região concentração de TAD, zoom na Figura 8	11

LISTA DE SIGLAS E ABREVIATURAS

ANAC	Agência Nacional de Aviação Civil
COMAER	Comando da Aeronáutica
DA	Defesa Aérea
INPE	Instituto Nacional de Pesquisas Espaciais
KML	Keyhole Mark-up Language
NM	Nautical Mile – 1,852 Km
PF	Polícia Federal
RFB	Receita Federal do Brasil
SISDABRA	Sistema de Defesa Aeroespacial Brasileiro
TAD	Tráfego Aéreo Desconhecido
TADREL	Relatório de Tráfego Aéreo desconhecido

LISTA DE SÍMBOLOS



Posição de primeira ou de última detecção



Posição estimada, também chamada de posição extrapolada



Pista de pouso e de decolagem – Aeródromo

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
2 PREPARAÇÃO DO TRABALHO.....	2
2.1. Análise da Base de Dados	2
2.2. Avaliação de viabilidade de mineração de dados.....	3
2.3. Operação com os atributos pré-processados	4
2.4. Estratégia adotada	5
2.5. Reconhecimento dos padrões	8
2.6. Trabalhos futuros	12
3 CONCLUSÃO	12
REFERÊNCIAS BIBLIOGRÁFICAS	14

1 INTRODUÇÃO

O COMAER é a Organização responsável por garantir a Soberania do Espaço Aéreo Brasileiro. Para que isso seja assegurado, inúmeras medidas são adotadas, sendo uma delas a efetiva identificação das aeronaves que utilizam o Espaço Aéreo Brasileiro. Essa identificação pode ser realizada através do Plano de Vôo apresentado, contato bilateral, via rádio, com os órgãos de controle, código *transponder*, dentre outros métodos.

Caso uma aeronave não seja adequadamente identificada, o SISDABRA é acionado e medidas adicionais são executadas, sendo uma delas a identificação visual por uma aeronave de DA, acionada sempre que a situação exigir. Caso o acionamento da aeronave de DA não seja requerido e o TAD não seja identificado através dos outros meios, ele será incluído em um relatório específico.

Uma das atribuições do SISDABRA é impedir o uso do espaço aéreo por aeronaves envolvidas com atividades ilícitas, sendo as principais o contrabando e o narcotráfico.

O objetivo de identificar regiões de origem e de destino, das aeronaves utilizadas para apoio a ações ilegais, norteou o trabalho realizado. Com uma vasta área de responsabilidade, a busca aleatória de rotas e de pistas de pouso torna-se quase que inviável e ineficiente. O desenvolvimento de uma ferramenta automatizada para a busca de padrões é uma alternativa plausível e viável, que agregará recursos visuais e estatísticos aos processos existentes.

No relatório de registro dos TAD, todos os dados pertinentes são cadastrados, sendo que, no presente trabalho, serão analisados apenas os dados pertinentes ao problema apresentado, as demais informações não fazem parte do escopo do problema.

2 PREPARAÇÃO DO TRABALHO

2.1. Análise da Base de Dados

A base de dados disponibilizada pelo COMAER é composta de uma tabela de 1.472 linhas por 32 colunas.

Cada linha ou instância da tabela representa a detecção de uma aeronave em um período de tempo. Os dados são obtidos durante o acompanhamento da aeronave, enquanto evolui dentro da cobertura radar do Espaço Aéreo Brasileiro.

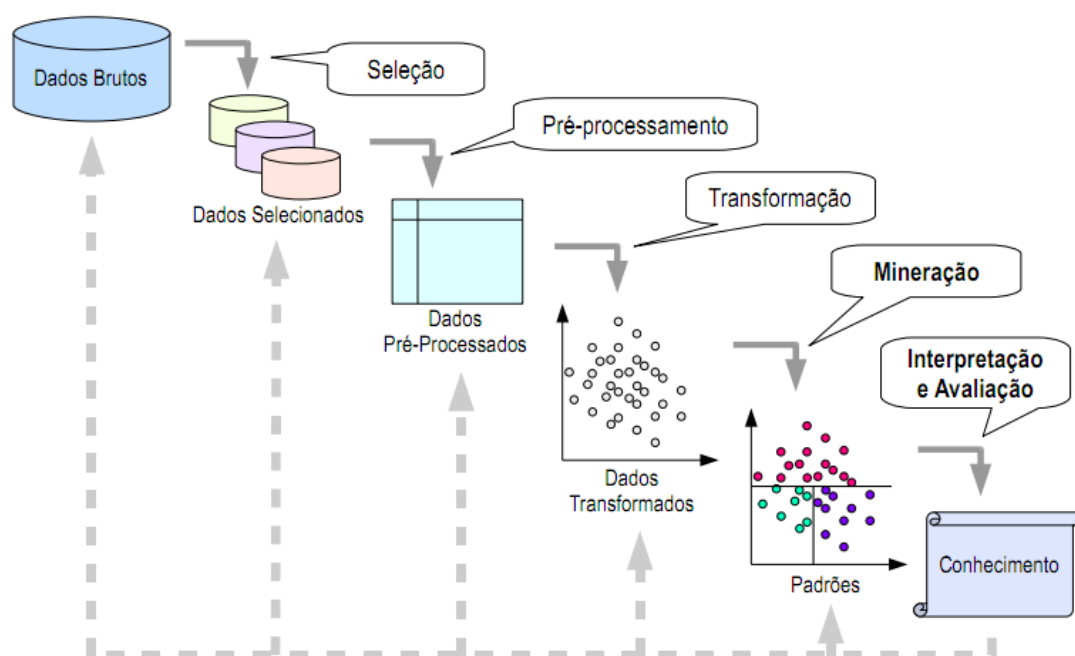


Figura 1 – Sequência do processo de mineração de dados

FONTE: Adaptada de Fayyad (1996).

Após análise da base de dados fornecida pelo COMAER, representada na Fig.1 pelos dados brutos, realizou-se uma seleção dos dados que são pertinentes ao objetivo de identificar possíveis aeródromos utilizados por aeronaves envolvidas com ilícitos. Dessa forma, elencaram-se oito atributos relevantes, dentre eles estão a posição geográfica (latitude e longitude) da

primeira vez que a aeronave foi detectada e da última detecção, além da data e do horário dessas observações.

2.1.1. Descrição dos atributos da tabela original, após a seleção

- a) Latitude: esse atributo é representado por um número inteiro de cinco dígitos na forma de sggmm (sinal, grau e minuto). Onde o sinal positivo representa o hemisfério Norte e o sinal negativo representa o hemisfério Sul. Ex.: a representação de -1020, observada na tabela original, significa 10° 20' S.
- b) Longitude: esse atributo é representado por um número inteiro de cinco dígitos na forma de sggmm (sinal, grau e minuto). Onde o sinal positivo representa o hemisfério oriental e o sinal negativo representa o hemisfério ocidental. Ex.: a representação de -5330, observada na tabela original, significa 53° 30' W.
- c) Data: o atributo data é representado por um número inteiro de 5 ou 6 dígitos na seguinte forma: ddmmaa ou dmmaa. Ex.: a representação 100510 equivale a seguinte data 10/05/2010. Da mesma forma, 50710 representa 05/07/2010.
- d) Hora: Representado na forma de um número inteiro na representação de 24 horas, conforme o formato hhmm. A hora segue o padrão internacional adotado pela aviação que é referenciado pelo meridiano de Greenwich (fuso Z). Ex.: uma observação realizada às 10:30 horário de Brasília será representada por 1330, devido a diferença de mais 3 horas entre o fuso P, de Brasília, e o fuso Z, Greenwich.

2.2. Avaliação de viabilidade de empregar a mineração de dados

Esta fase caracteriza o pré-processamento das informações e objetiva verificar, preliminarmente, a existência de padrões nas informações contidas na base de dados. Utilizou-se um gerador de arquivo .kml, disponível na internet em [1],

para permitir uma pré-análise da existência de convergência ou divergência nas rotas observadas. Os dados foram manipulados manualmente com uma amostra dos dados, porém com significativa possibilidade de sucesso. A figura 2 representa algumas dessas rotas, onde pode ser observada uma pista de pouso próxima, ao local onde foi realizada a última detecção. Nesses locais normalmente não dispõe de um órgão do sistema de controle de tráfego aéreo, logo as aeronaves evoluindo nessas direções não são obrigadas a identificarem-se. Dessa forma, são classificadas como desconhecidas e inseridas na base de dados. Vale salientar que, com base nas informações obtidas, não é possível inferir que as aeronaves observadas estejam envolvidas com atividades ilícitas. Para essas conclusões serem validas, será necessário a confrontação das observações realizadas, neste trabalho, com outras informações de inteligência. Também, não é possível afirmar que a aeronave pousou ou decolou de uma determinada pista, apenas existe indício disso.

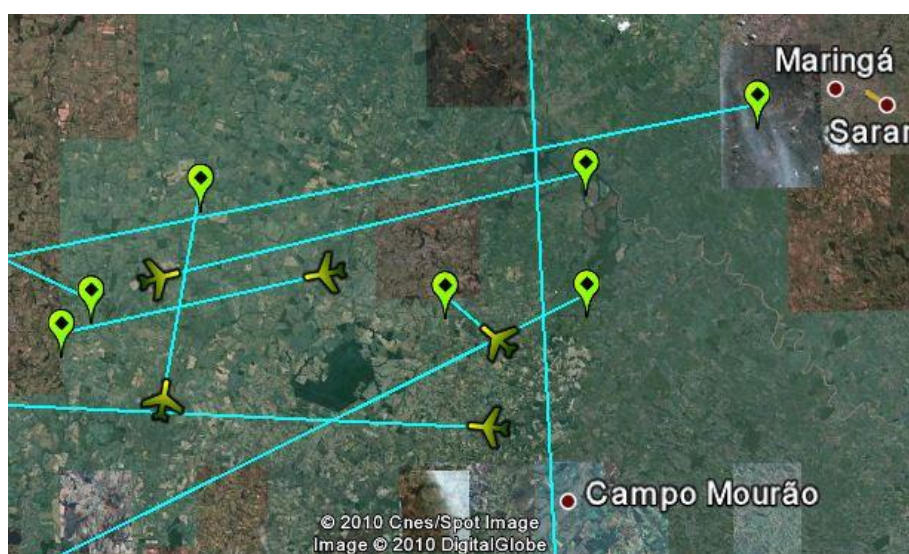


Figura 2 – Convergência de 4 rotas para Maringá

2.3. Operação com os atributos pré-processados

Após a análise de viabilidade ter demonstrado a possibilidade de sucesso no processo de análise das rotas, realiza-se a seleção dos atributos e de todas as

instâncias que são efetivamente utilizados. Com isso, uma planilha do Microsoft Office Excel 2007 com a extensão .xlsx é gerada.

A operação manual em toda a BD não é viável, para isso uma aplicação especializada é necessária. A plataforma JAVA demonstra grande flexibilidade e possui API, já elaborada, que otimiza o processo. Para que o aplicativo JAVA seja viável, inicialmente a planilha .xlsx deve ser exportada para o padrão .csv (separado com vírgula). A adequação do formato existente para um formato compatível com a aplicação é necessária. Esse processo caracteriza a fase de transformação da Figura 1. Assim, as latitudes e longitudes são modificadas para um valor Double, no formato de gg,gg (graus e décimos de grau). Dessa forma, a representação de 23°30'W passa a ser -23,5 graus. Também, transformar hora e minutos para valor um valor inteiro equivalente a minutos.

Para auxiliar na identificação das informações há a necessidade dos dados serem transformados conforme figura 1. A aplicação desenvolvida disponibiliza uma tabela, contendo o total de pontos em cada região pré-definida, e dois arquivos .kml, um expõe os pontos e rotas e o outro delimita as áreas de interesse, ambos arquivos .kml são visualizados no Google Earth.

2.4. Estratégia adotada

Para efeito de avaliação de rotas utilizadas por aeronaves de interesse, o horário ou o dia em que elas podem ocorrer não é relevante, pois as características das aeronaves utilizadas são muito suscetíveis a meteorologia, restando a elas operarem quando houver condições favoráveis.

A adoção dos atributos de proa e velocidade, no momento da última detecção, disponíveis na BD original, não serão utilizados, mas sim o rumo médio e velocidade mantidos entre os pontos iniciais e finais da observação.

Ainda com relação ao processo de mineração, adaptado de Fayyad, figura 1, deve-se iniciar a mineração dos dados propriamente dita, buscando identificar a existência de padrões entre eles.

Observa-se na representação do Google Earth que os pontos de início e de término da detecção não são coincidentes com pistas de pouso ou decolagem. Isso se deve ao fato que a aeronave estará em área de cobertura radar, a partir de uma determinada altura e de certa distância de um radar. Para resolver a lacuna existente entre o ponto de origem e a primeira detecção, assim como, ponto de última detecção e o destino foi estipulado à criação de posições futuras estimadas da aeronave (extrapolações), supondo que ela estivesse ainda sendo acompanhada pelo radar e que ela mantivesse as características médias de vôo (rumo e velocidade). Para esse estudo, as extrapolações foram limitadas em 10 minutos antes e depois das detecções iniciais e finais. Tanto as posições iniciais, finais, as extrapolações e uma linha unindo esses pontos é gerado em um arquivo chamado KmlTAD.kml.

As regiões de agrupamento serão definidas, inicialmente, pelas latitudes e longitudes em graus inteiros (10°S, 11°S ...). Caso as regiões de agrupamento possuam um valor significativos de amostras, valor esse definido pelo usuário, o aplicativo gera um outro arquivo, denominado Quadricula.kml. Esse arquivo contém os limites das quadrículas mais representativas. Caso as amostras ultrapassem novo limite, será gerada uma subdivisão de 50% (10°S, 10,5°S, 11°S...).

A delimitação de uma região é necessária para restringir uma área de busca, tendo em vista que as posições são estimadas não tendo como definir um ponto geográfico específico. Na prática as rotas não levam necessariamente até a vertical da pista, pois são necessários procedimentos, comuns nas decolagens e nos pousos, que levam o piloto, inicialmente, a realizar deslocamentos em direções que não coincidem com a sua rota. Geralmente esses procedimentos são para separação de obstáculos, separação com

outros tráfegos, fatores meteorológicos, conforto do piloto ou procedimentos específicos da aeronave. Dessa forma, o comportamento da aeronave fora da cobertura radar torna-se aleatório, mais uma razão para a adoção de área de busca, ao invés de ponto geográfico.

A figura 3 representa uma rota hipotética, onde podem ser observados os pontos de primeira e de última detecção, além das posições extrapoladas, isto é, sem a detecção da aeronave realizada por um radar. Essa rota representa uma aeronave decolando de uma pista não controlada, nas proximidades de Curitiba-PR, com destino a outra pista não controlada, nas proximidades de São Paulo-SP. Durante o vôo, a aeronave foi detectada por um radar e, devido às suas características de vôo, não foi considerada uma ameaça a Segurança Nacional, dessa forma, somente será realizado o registro no relatório de TAD. As quadrículas de número 1, 2 e 4, por apresentarem reduzido número de pontos ($\leq 5^1$) estão na escala de 60NM (1° de latitude), já a quadrícula de número 3 apresenta 6 pontos, logo ela foi novamente dividida, passando cada nova quadrícula a escala de 30NM (reduzida em 50%).



Figura 3 – Rota simulada entre Curitiba e São Paulo

¹ O valor de 5 pontos foi atribuído, arbitrariamente, para exemplificação. Para emprego normal, o valor deve ser adequado às necessidades do usuário.

2.5. Reconhecimento dos padrões

Devido ao caráter sigiloso da BD, algumas informações serão omitidas nesse relatório, porém nada que comprometa o entendimento do processo executado para atingir o objetivo proposto.

Após a execução do aplicativo em Java e obtida uma tabela contendo os agrupamentos, figura 4, onde se percebe, claramente, a existência de clusters. As áreas (latitudes e longitudes), as quais se referem esses valores foram propositalmente omitidas.

139	216	46	34	324	30
162	190	5	50	314	102
95	450	912	353	263	59
44	68	512	35	2	4
22	76	31	22	0	1
20	1	6	1	11	0
0	1	2	0	0	0
3	3	0	0	0	1

Figura 4 – Tabela de agrupamentos, contendo posição inicial, final e extrapolações

A simples identificação na tabela, figura 4, dos setores onde há uma maior concentração de pontos não é suficiente para identificar locais que possam ser utilizados para pouso e decolagem de aeronaves. Para isso esses setores devem ser transferidos a um mapa, onde seja viabilizada uma procura manual por esses locais. Nesse trabalho foi utilizada a plataforma do Google Earth.

O conteúdo gerado pelo arquivo Quadricula.kml, representando as áreas com maior concentração de pontos, pode ser observado na Figura 5. Havendo, entre 450 e 800 pontos ou amostras, a representação é realizada na escala de

1° de latitude (60 NM), como em 1 e 2, na Figura 5. Caso sejam mais de 800 pontos a representação da quadrícula é na escala de 0,5° de latitude (30NM), como em 3, na figura 5.

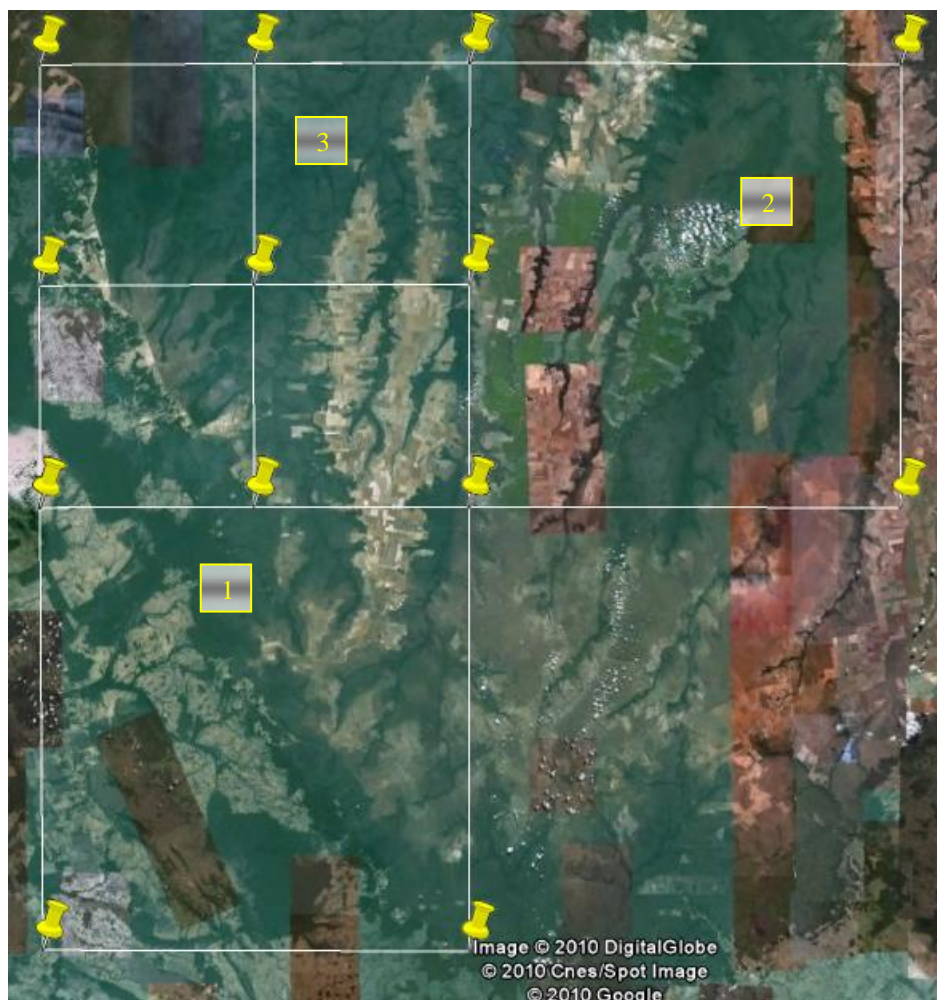


Figura 5 – Quadrículas de 60NM e de 30NM

Para realizar a busca por locais de pouso e decolagem é necessário que, além das áreas de interesse da figura 5, haja a representação das rotas voadas e as extrapolações implementadas. Assim, é oportuno realizar a superposição das informações contidas no arquivo KmlTAD.kml com o Quadrícula.kml.



Figura 6 – Visualização de quadrícula e agrupamento de rotas



Figura 7 – aproximação da Fig 6, onde são observadas 2 pistas de pouso.

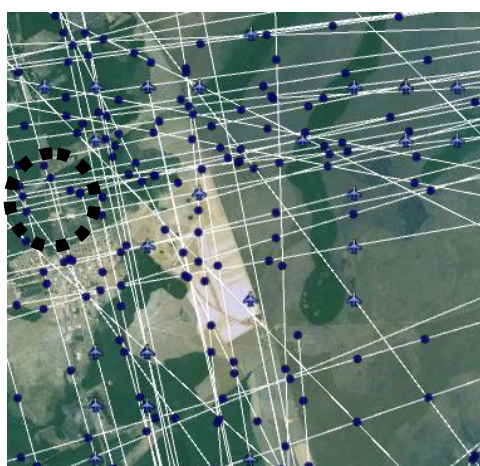


Figura 8 – Superposição de rotas e quadrículas

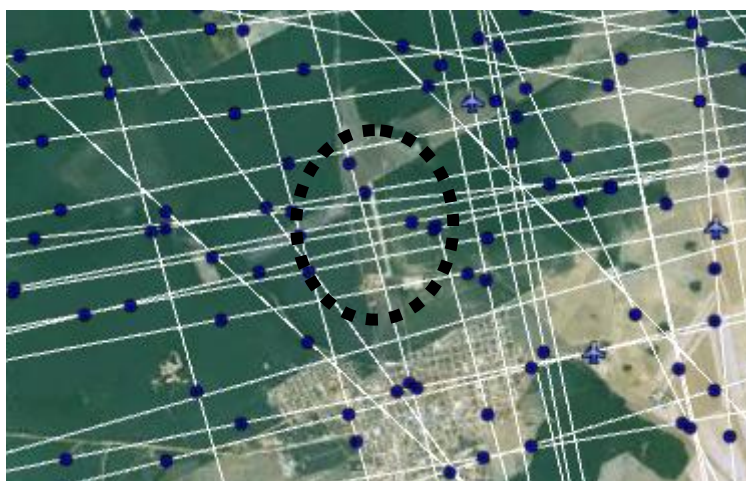


Figura 9 – Procura manual por pista de pouso na Figura 8



Figura 10 – Pista de pouso localizada em região concentração de TAD, zoom na Figura 8

Como pode ser observado nas figuras 6, 8 e 9, existe uma grande concentração de pontos de início ou finais, sendo indícios de que as aeronaves possam ter decolado ou pousado naquela região. A medida que é realizada uma busca visual na região demarcada é possível localizar várias pistas como

nas figuras 7, 9 e 10. Porém, não é cabível afirmar que qualquer das aeronaves estava cometendo irregularidades, tão pouco, ilegalidades. Para isso, é necessário haver a combinação dessas informações com outras de inteligência.

2.6. Trabalhos futuros

O emprego da metodologia acima se mostrou bastante adequada para identificar áreas de decolagem e de pouso de aeronaves desconhecidas. A otimização da busca desses locais é crucial para medidas mais efetivas de averiguação das aeronaves pelos organismos de fiscalização (COMAER, PF, RFB e ANAC). Para isso, a automatização da identificação desses locais é desejável.

A modificação do sistema de registro dos TAD, contemplando todas as posições intermediárias, poderá agregar mais confiabilidade a BD atual, pois as diferenças temporais e espaciais em alguns casos é muito significativa e o emprego da média de rumo e de velocidade pode induzir a erros.

3 CONCLUSÃO

A adoção da metodologia apresentada proporciona a identificação de regiões onde há uma maior incidência de vôos não identificados. Baseado nas constatações obtidas nesse trabalho, haverá uma maior eficiência no emprego de operações rotineiras de policiamento do espaço aéreo, empregando os dados coletados para planejar ações localizadas.

O emprego de técnica de mineração de dados, com a clusterização de rotas de tráfegos aéreos desconhecidos, visando identificar os possíveis locais de decolagem ou de pouso desses tráfegos, demonstrou-se muito adequada.

A associação de tabelas e de visualização, no Google Earth, permite uma maior clareza quanto a existência de clusters de pontos que induzem a locais de pouso. Com isso a busca manual para localizar essas pistas é bastante facilitada.

A adoção dessa metodologia pode servir de embrião para ampliar o escopo do projeto para a busca automática de pistas de pouso.

REFERÊNCIAS BIBLIOGRÁFICAS

COMANDO DE DEFESA AEROESPACIAL BRASILEIRO (COMDABRA). Base de Dados de Relatório de Tráfego Aéreo Desconhecido – RESERVADO, obtido através de malote oficial do COMAER, através de CD-ROM.

Fayyad, U, Piatetsky-Shapiro, G., Smyth, P., Knowledge Discovery and Data Mining: Towards a Unifying Framework, in: Proc. KDD-96, Portland, OR, AAAI Press, Menlo Park, CA, 1996, pp. 82–88.

GOOGLE EARTH. **Tools for Google Earth**. Disponível em: <http://www.earthpoint.us/ExcelToKml.aspx>. Acesso em 31 out 2010.

GOOGLE GROUP. **KML Framework**. Disponível em: <http://code.google.com/p/kmlframework/>. Acesso em 31 de outubro de 2010.

Fayyad, U, Piatetsky-Shapiro, G., Smyth, P., Knowledge Discovery and Data Mining: Towards a Unifying Framework, in: Proc. KDD-96, Portland, OR, AAAI Press, Menlo Park, CA, 1996, pp. 82–88.