



Programa Interdisciplinar de Pós-Graduação em
Computação Aplicada
Mestrado Acadêmico

Róbson Koszeniewski Gomes

Uma Abordagem para Detecção Automática de Planos em Modelos
Digitais de Afloramentos Baseada em PCA

São Leopoldo, 2014

Róbson Koszeniewski Gomes

UMA ABORDAGEM PARA DETECÇÃO AUTOMÁTICA DE PLANOS EM MODELOS
DIGITAIS DE AFLORAMENTOS BASEADA EM PCA

Orientador:
Prof. Dr. Luiz Gonzaga da Silveira Junior

Co-orientador:
Prof. Dr. Luiz Paulo Luna de Oliveira

São Leopoldo
2014

G633a

Gomes, Róbson Koszeniewski.

Uma Abordagem para Detecção Automática de Planos em Modelos Digitais de Afloramentos Baseada em PCA / Róbson Koszeniewski Gomes. São Leopoldo, 2014.

68 f.: il. ; 30cm.

Orientador: Prof. Dr. Luiz Gonzaga da Silveira Junior.

Dissertação (Mestrado) — Universidade do Vale do Rio dos Sinos, Programa Interdisciplinar de Pós-Graduação em Computação Aplicada – PIPCA, São Leopoldo, 2014.

1. Modelagem Digital de Afloramentos. 2. Análise de Componentes Principais. I. Título

CDU 004.4

Ficha catalográfica elaborada por Sara Caselani Zilio – CRB 10/1695

Deus.

For God so loved the world, that he gave his only begotten Son, that whosoever believes in him should not perish, but have everlasting life. — JOHN 3:16

AGRADECIMENTOS

Primeiramente gostaria de agradecer a Deus por mais uma vez me abençoar com amigos que me auxiliaram fortemente até o final desta jornada. Agradeço também minha família que, mais uma vez, teve que compreender a minha ausência e mau humor algumas vezes. Minha ex-namorada, agora noiva, Emili, pelo auxílio na revisão e palavras de calma.

Agradeço aos meus grandes amigos Lucas e Fabrício que diversas vezes tiveram a calma e paciência, me compreendendo e auxiliando nos momentos mais complicados. Juntos conquistamos mais este desafio, e juntos saímos vitoriosos. Ao meu amigo Berni que sempre está online para aquelas partidas casuais de diversos jogos, e que com as quais pude me divertir após algumas horas de estudo.

Devo boa parte desta vitória aos professores que estiveram ao meu lado desde o primeiro dia de aula. Luiz Gonzaga meu orientador foi paciente durante a construção do trabalho devido a atrasos ocorridos, e rígido nos momentos necessários. Ao professor Luna que deu grandes idéias para que o trabalho tivesse a qualidade necessária de um trabalho de mestrado. Aos professores da geologia Maurício e Chico que me derão todo apoio nas questões técnicas da área. Agradeço também ao colega Marcelo que cedeu toda a base do seu trabalho, assim como algumas horas de conversa.

Também tenho que agradecer a empresa em que hoje trabalho, PROCERGS, pela grande ajuda financeira, a qual sem, este desafio nunca teria sido conquistado. Aos colegas de trabalho Israel e James que me auxiliaram com dicas de formatação de diversas partes do texto.

Enfim gostaria de agradecer a todos que de certa forma me ajudaram, direta ou indiretamente, para a concretização deste trabalho. Com quem tive a conversa repetitiva sobre o trabalho, ou conversas para esquecer dos problemas, todos influenciaram neste trabalho e como ele foi feito. A todos muito obrigado!

RESUMO

A coleta de dados espaciais tem sido intensamente empregada na área geológica, através da técnica de LIDAR (*Light Detection and Ranging*). Este tipo de sensoriamento digital remoto de alta resolução e precisão, resulta em modelos digitais 3D que permitem uma análise mais detalhada e quantitativa de estruturas heterôgeneas, como afloramentos. Um dos estudos realizados pelos geólogos são análises sobre a geometria da formação de rochas, onde a informação de orientação de um plano inclinado é um indicativo para a compreensão global da estrutura. Este trabalho propõe a utilização da técnica de Análise de Componentes Principais (PCA) para calcular e detectar automaticamente todos os planos em uma nuvem de pontos. Uma ferramenta foi construída para implementar a visualização do modelo digital e apurar os melhores planos. Um estudo foi realizado a fim de validar as informações encontradas pelo método proposto e dados medidos em campo.

Palavras-chave: Afloramentos. LIDAR. Análise de Componentes Principais. Ângulo de Mergulho e Direção da Camada.

ABSTRACT

The use of LIDAR (Light Detection and Ranging) systems for gathering spatial data has been extensively used in geological studies. This type of digital remote sensing delivers high resolution and accuracy, resulting in 3D digital models which allow a more detailed and quantitative analysis of heterogeneous structures, as outcrops. One of the studies is based on analysis of the forming rocks geometry. The orientation of a slope plane is an indication for the overall understanding of the structure. This work proposes a new method to automatically compute and detect all possible planes in a point cloud, based on Principal Component Analysis (PCA) technique. A software tool was constructed to implement the digital model visualization and compute the best planes. A study was conducted to compare and validate the results of the method and the field data collected.

Keywords: Outcrop. LIDAR. Principal Component Analysis. Dip and Strike.

LISTA DE FIGURAS

Figura 1:	Afloramento	23
Figura 2:	Medindo a direção da camada de um plano	24
Figura 3:	Medindo o ângulo de mergulho de um plano	25
Figura 4:	Direção da camada e ângulo de mergulho do plano	25
Figura 5:	LASER Scanner Terrestre	26
Figura 6:	Aplicações do <i>LASER Scanner</i>	27
Figura 7:	LIDAR	28
Figura 8:	Nuvem de pontos afloramento	29
Figura 9:	Plano PCA	40
Figura 10:	Avaliação do plano	41
Figura 11:	Valores de ajuste e confiabilidade dos planos	42
Figura 12:	Algoritmo de detecção automática de planos	43
Figura 13:	Divisão dos pontos por grupos	44
Figura 14:	Grafo de cena	47
Figura 15:	Função Sigmoidal	48
Figura 16:	Plot e renderização da superfície sintética	51
Figura 17:	Testes realizados alterando o parâmetro de ajuste do plano m	52
Figura 18:	Equipamento ILRIS3D da Optech	53
Figura 19:	Afloramento pedreira Incopel	54
Figura 20:	Plano de estudo	55
Figura 21:	Análise dos pontos que formam o plano	56
Figura 22:	Afloramento pedreira Incopel - corte	57
Figura 23:	Planos detectados sobre o afloramento	58
Figura 24:	Grandes planos no afloramento para $m \geq 5$	59
Figura 25:	Planos no afloramento para $m \geq 6$	60

LISTA DE TABELAS

Tabela 1:	Medições de campo	54
Tabela 2:	Dados de orientação do plano - isolado	56
Tabela 3:	Diferenças dos dados de orientação do plano - isolado	56
Tabela 4:	Dados de orientação do plano - corte	57
Tabela 5:	Diferença dos dados de orientação do plano - corte	57
Tabela 6:	Dados de orientação do plano - afloramento	58
Tabela 7:	Diferenças dos dados de orientação do plano - afloramento	59
Tabela 8:	Dados dos grandes planos no afloramento para $m \geq 5$	60

LISTA DE SIGLAS

PCA	Principal Component Analysis
LiDAR	Light Detection and Ranging
TLS	Terrestrial LASER Scanner
DD	Dip Direction
DA	Dip Angle
LEV	Log-Eigenvalue
OSG	Open Scene Graph
LGPL	GNU Lesser General Public License
LASER	Light Amplification by Stimulated Emission

SUMÁRIO

1 INTRODUÇÃO	19
1.1 Motivação	20
1.2 Objetivo	20
1.3 Organização do texto	21
2 CONCEITOS BÁSICOS	23
2.1 Afloramento	23
2.1.1 Direção da camada (DD) e Ângulo de Mergulho (DA)	24
2.2 LASER Scanner Terrestre (TLS)	26
2.2.1 Nuvem de Pontos	29
2.2.2 Atributos de orientação de planos	30
2.3 Análise de Componentes Principais (PCA)	30
2.3.1 Planos com PCA	33
3 TRABALHOS RELACIONADOS	35
3.1 Detecção de planos	35
3.2 Análises geológicas automáticas	36
4 DETECÇÃO AUTOMÁTICA DE PLANOS EM AFLORAMENTOS DIGITAIS	39
4.1 Calculando o plano de melhor ajuste	39
4.2 Detecção automática de planos	42
4.2.1 Divisão dos dados de entrada	43
4.3 Implementação	45
4.3.1 Detalhes do aplicativo	45
4.3.2 Função de cor dos planos	46
5 RESULTADOS	51
5.1 Superfície sintética	51
5.2 Modelo digital de afloramento	53
5.2.1 Primeiro teste: plano isolado	55
5.2.2 Segundo teste: corte no modelo	56
5.2.3 Terceiro teste: modelo completo	57
6 CONCLUSÃO	61
REFERÊNCIAS	63

1 INTRODUÇÃO

Afloramentos são locais onde as rochas que formam a crosta terrestre ficam expostas. Isso ocorre devido a erosões naturais ou ações humanas. Esta é uma fonte básica de pesquisa para geólogos (GROTZINGER et al., 2007). Uma análise comum realizada em geologia é a de medição de estruturas planares em afloramentos. Este tipo de pesquisa objetiva a construção de mapas estruturais para determinar a composição e formação das rochas.

As estruturas planares de um afloramento são definidas de acordo com os valores de orientação do plano. Estas medidas são o ângulo de mergulho (*DA - Dip Angle*) e o direção da camada (*DD - Dip Direction*).

Direção da camada é definido como a direção do compasso de uma camada de rocha referente a sua intersecção com a superfície horizontal. Ângulo de mergulho é medida em uma linha perpendicular a direção da camada, e indica a inclinação da camada de rocha (GROTZINGER et al., 2007; DAVIS; REYNOLDS, 1996).

O método convencional que estas medições são realizadas são com equipamentos como bússola e clinômetro. Em afloramentos de difícil acesso, ou onde existe interferência eletromagnética (depósitos de minerais magnéticos), métodos tradicionais podem levar a erros. Isto torna difícil e se os erros não forem corrigidos podem provocar análises de estruturas geológicas equivocadas. Entretanto, é necessário coletar e processar dados para o entendimento da estrutura do afloramento.

Para resolver tais situações, vem sendo intensamente empregado o uso de técnicas de sensoriamento digital remoto. Como resultado são obtidos modelos digitais 3D que permitem uma análise mais detalhada e quantitativa (BUCKLEY et al., 2008).

Uma das técnicas aplicadas é a utilização de *LASER scanner* terrestres (*TLS - Terrestrial LASER Scanner*). TLS é um equipamento que realiza uma amostragem de objetos alvo com alta densidade em um curto espaço de tempo. Como resultado do processo de mapeamento do afloramento, um modelo digital é constituído e utilizado para análise e interpretação dos dados de campo (BUCKLEY et al., 2008).

O modelo digital de afloramento é um conjunto de pontos que possuem coordenadas tridimensionais conhecidas (nuvem de pontos). Aplicando técnicas matemáticas é possível calcular o plano aproximado a partir de um conjunto de dados. Análise de componentes principais (*PCA - Principal Component Analysis*) é uma técnica de análise de dados com o objetivo de reduzir a dimensionalidade de um conjunto de dados com uma grande quantidade de variáveis relacionadas, preservando a variação original dos dados (JACKSON, 2005).

Com PCA é possível analisar o modelo digital do afloramento para detectar as estruturas planares estudadas em campo. Sabendo a posição dos planos virtuais e seu vetor normal é possível calcular os valores de orientação do plano (ALAN K. L. KWONG; WONG, 2007).

1.1 Motivação

Na natureza existem afloramentos com quilômetros de extensão e dezenas de metros de altura. Nestes casos, pesquisas de campo sofrem restrições tais como: tempo de coleta de dados em campo, condições climáticas e acessibilidade. Análises mais arriscadas são evitadas pois levam perigo ao pesquisador. Mesmo em locais com acessibilidade, por motivos geológicos, as medições podem ser imprecisas ou invalidadas.

Utilizando equipamentos de sensoriamento remoto como o TLS estas restrições são removidas (BUCKLEY et al., 2010). Obtendo o modelo digital é possível realizar as mesmas análises de campo em ambientes fechados como escritórios. Neste caso outras ferramentas devem ser utilizadas. O suporte de técnicas de visualização facilitam o estudo e processamento destes dados.

É possível calcular os valores de orientação dos planos em afloramentos através da nuvem de pontos do modelo. Inicialmente os pontos referentes ao plano devem ser identificados manualmente. Com a seleção um plano é gerado e as informações de orientação são calculadas.

Utilizando as características da técnica PCA é possível que as estruturas planares do modelo digital de afloramento possam ser detectadas automaticamente. Este processo automatizado trará avanços para a análise dos dados obtidos com o mapeamento remoto de estruturas geológicas como os afloramentos.

1.2 Objetivo

Com a possibilidade de detecção e cálculo de estruturas planares utilizando o modelo digital de afloramentos o presente trabalho tem como objetivo principal: Detectar automaticamente planos e extrair as informações de direção da camada e ângulo de mergulho de cada plano a partir dos dados do TLS.

Assim se destacam os seguintes objetivos específicos:

- a) analisar o método de componentes principais e aplica-lo para detecção dos planos geométricos;
- b) criar um algoritmo para detecção automática dos planos com base nos dados do modelo 3D de afloramentos;
- c) implementar uma ferramenta para visualização da nuvem de pontos, dos planos detectados e as informações de orientação do plano;
- d) analisar os resultados obtidos comparando com as medições realizadas em campo, quanto a detecção dos planos e informações de orientação dos mesmos.

1.3 Organização do texto

Este trabalho está organizado da seguinte maneira: o Capítulo 2 apresenta conceitos sobre os afloramentos e estudos de estruturas planares em campo. É abordado a utilização do *LASER scanner* terrestre para o mapeamento do afloramento e como os dados estudados em campo podem ser obtidos a partir do modelo digital. Ainda no Capítulo 2 a técnica de Análise de Componentes Principais será apresentada, e como ela é utilizada neste trabalho para calcular os planos de melhor ajuste em uma nuvem de pontos.

No Capítulo 3 são brevemente apresentados alguns trabalhos relacionados tanto na área de detecção de planos como também em análise de outros aspectos geológicos pertinentes.

No Capítulo 4 é apresentado em detalhes como um plano é calculado utilizando PCA para uma nuvem de pontos, assim como o método proposto para detecção automática de diversos planos em um modelo digital de afloramentos. A implementação para visualização dos resultados é descrita em detalhes, citando os parâmetros criados e a função parametrizada de cor para os planos detectados. No Capítulo 5 os resultados são apresentados e a comparação com os dados reais de um afloramento, obtidas utilizando o TLS, são detalhados.

No Capítulo 6 as considerações finais são colocadas, incluindo os trabalhos futuros visando melhorias no processo de detecção automática de planos.

2 CONCEITOS BÁSICOS

Neste capítulo serão apresentados os conceitos básicos utilizados no decorrer deste trabalho. Uma das principais fontes de estudo de geólogos, o afloramento, será brevemente descrito na Seção 2.1. Assim como as estruturas planares encontradas nos afloramentos. Como tais estruturas são medidas em campo é descrito na Seção 2.1.1. A Seção 2.2 introduz o equipamento TLS que é utilizado para geração de nuvem de pontos com coordenadas 3D de um afloramento. O resultado deste mapeamento é uma nuvem de pontos que forma o modelo digital do afloramento, descrito na Seção 2.2.1. Na Seção 2.2.2 é apresentado como os dados de orientação de estruturas planares de afloramentos podem ser calculados a partir da nuvem de pontos. Por fim a Seção 2.3 descreve o método de Análise de Componentes Principais, que é utilizado neste trabalho para calcular os planos de melhor aproximação na nuvem de pontos.

2.1 Afloramento

A crosta terrestre é composta de rochas, contudo, os tipos de rochas diferem na maioria dos lugares. Em locais onde as rochas estão expostas, são chamadas de afloramentos. Afloramentos são excelentes fontes de estudo para geólogos, pois permite determinar a composição local das rochas (GROTZINGER et al., 2007).

Afloramentos não cobrem a maior parte da superfície terrestre, pois em muitos casos a rocha está coberta com solo e vegetação. Em locais onde ocorrem erosões ou movimento da crosta por placas tectônicas, a rocha pode ficar exposta, formando o afloramento. A exposição da rocha tende a ocorrer em áreas onde a erosão é maior do que o crescimento da vegetação e efeitos do tempo, ou seja, em locais com desmoronamentos, montanhas ou áreas com placas tectônicas ativas. Afloramentos podem ser criados também devido a ações do homem como escavações, construção de rodovias, e outros. A Figura 1 ilustra um afloramento originado da escavação de uma pedreira.

Figura 1: Foto de um afloramento em uma pedreira, com uma altura aproximada de 3 metros.



Fonte: Imagem cedida por LASERCA/UNISINOS

Afloramentos permitem um estudo com contato direto e amostragem da rocha para criação do mapa geológico. Estudos diretos são cruciais para análise da formação da crosta terrestre, e afloramentos são extremamente importantes para estes estudos (JACKSON; MEHL; NEUENDORF, 2005).

Algumas informações que podem ser obtidas através do estudo direto de afloramentos são: características de orientação da geologia estrutural (fraturas, dobras, planos, etc.), características de depósito e orientação paleomagnética.

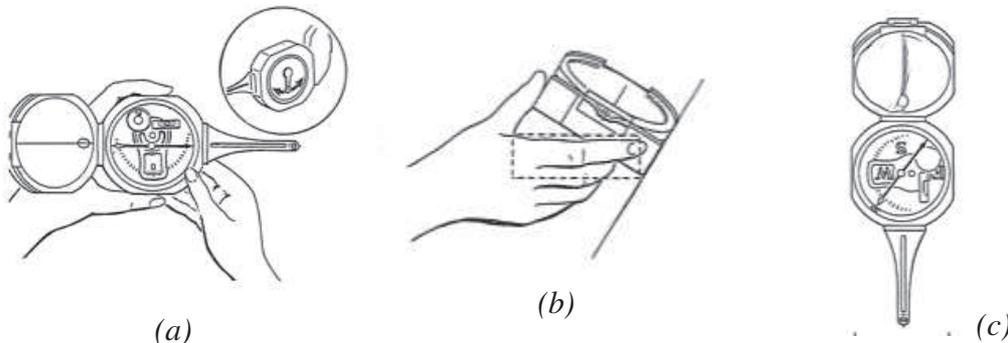
As medidas estruturais de planos são o foco deste trabalho. Este tipo de pesquisa identifica estruturas planares nos afloramentos. Isso permite que o geólogo possa construir mapas da região para identificação da formação local. Os planos medidos são classificados de acordo com sua orientação, definidos pelos valores de direção da camada (*Dip Direction* ou *Strike*) e ângulo de mergulho (*Dip Angle* ou *Dip*). O significado destes valores e como são obtidos no afloramento será detalhado na Seção 2.1.1.

2.1.1 Direção da camada (DD) e Ângulo de Mergulho (DA)

Durante as pesquisas de campo em afloramentos, geólogos realizam medições de estruturas planares. A identificação destes planos se dá através de duas medidas de orientação: direção da camada (*Dip Direction* ou DD) e ângulo de mergulho (*Dip Angle* ou DA).

A direção da camada é a medida que indica a relação entre o horizonte e a camada de rocha. A Figura 2 ilustra alguns passos para medir o valor de um plano utilizando um compasso com clinômetro. Inicialmente o clinômetro é configurado em zero graus para que seja utilizado como um nível (Figura 2-(a)). Posiciona-se um dos lados do compasso contra o plano (Figura 2-(b)), rotacionando o compasso até que a bolha fique nivelada. Para medir o ângulo formado, incline o compasso mantendo-o nivelado (Figura 2-(c)) para apurar o valor.

Figura 2: Medindo a direção da camada de um plano.

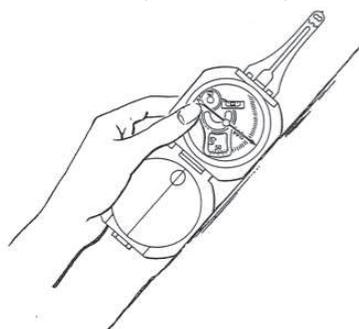


Fonte: Imagem retirada de (DAVIS; REYNOLDS, 1996)

De modo geral a direção da camada indica a posição da camada da rocha, onde o ângulo medido pela bússola é o azimuth. A Figura 4 ilustra a direção da camada e a linha de rumo de mergulho que é utilizada para medir o ângulo de mergulho.

O ângulo de mergulho do plano é medido com uma linha perpendicular a linha da direção da camada e indica os graus de inclinação com um plano horizontal (GROTZINGER et al., 2007; DAVIS; REYNOLDS, 1996). Para realizar a medida deve-se posicionar o compasso na face do plano e então rotacionar o clinômetro até que esteja nivelado (Figura 3).

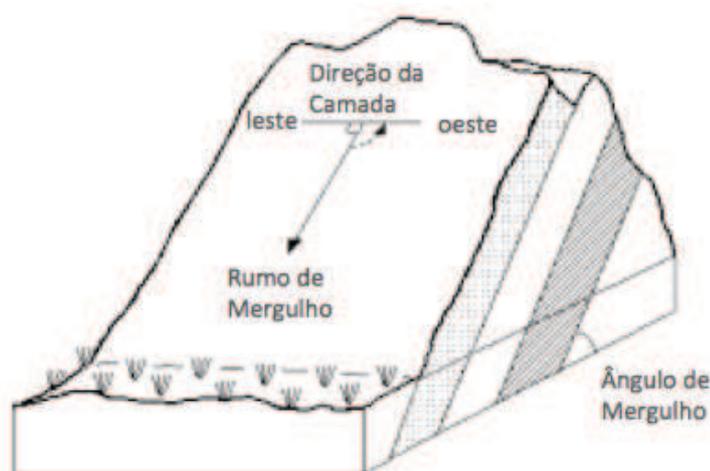
Figura 3: Medindo o ângulo de mergulho de um plano.



Fonte: Imagem retirada de (DAVIS; REYNOLDS, 1996)

A Figura 4 ilustra o ângulo de mergulho, criado medindo-se a inclinação da rocha na linha de sentido de mergulho.

Figura 4: Direção da camada e ângulo de mergulho do plano



O processo de medição dos valores de orientação do plano é realizado manualmente e varia de acordo com a experiência do especialista (SOUZA, 2012). A forma do afloramento e a altura das fraturas planares expostas podem comprometer o processo, tornando inviáveis e perigosas tais medições. Mesmo afloramentos de fácil acesso podem ser compostos de minerais magnéticos que desorientam a agulha do compasso.

Para solucionar este problema, técnicas de sensoriamento remoto vem sendo empregadas na área geológica. Equipamentos como o *LASER scanner* terrestre são utilizados para criar modelos digitais dos afloramentos.

2.2 LASER Scanner Terrestre (TLS)

LASER Scanner Terrestre é um equipamento que permite a amostragem de alta densidade em um curto espaço de tempo de objetos. Um objeto real é amostrado para coleta de dados sobre sua forma e, se utilizado um sensor de imagem como uma câmera digital, possivelmente aparência (cor). Os dados coletados podem ser utilizados para construção de um modelo digital, úteis para diversas aplicações.

Figura 5: LASER Scanner Terrestre Optech ILRIS3D com equipamento de georreferenciamento acoplado.



Fonte: Imagem cedida por LASERCA/UNISINOS

A utilização de LASER scanner terrestre (Figura 5) tem se tornado cada vez mais frequente devido a sua robusta abordagem para coleta de dados. É uma técnica efetiva de captura de dados digitais de alvos tais como afloramentos. Isto permite que as informações sejam visualizadas e analisadas de forma mais detalhada, removendo as restrições impostas por estudos de campo tais como:

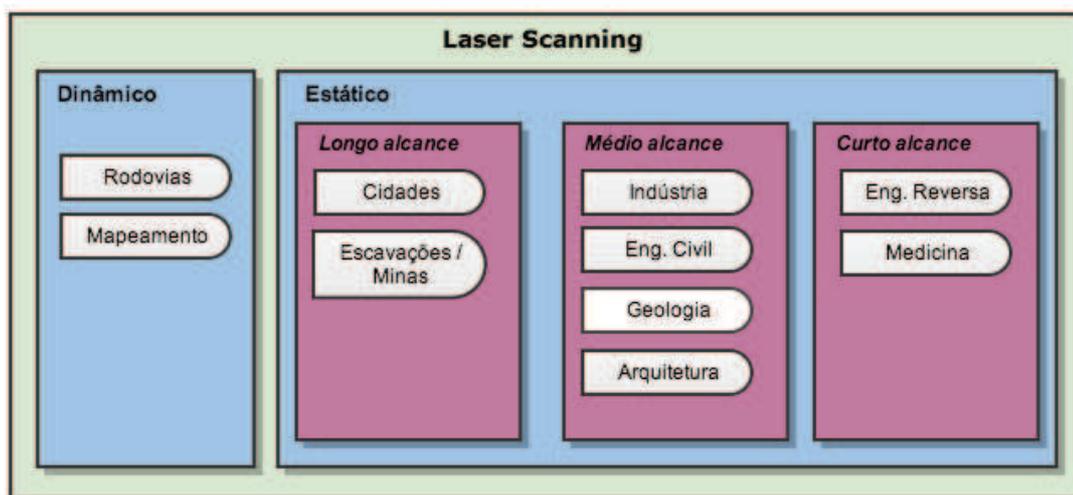
- **Limites de tempo:** usualmente estudos de campo estão limitados ao tempo disponível pelos profissionais. Há locais de pesquisa afastados de áreas urbanas que exigem um acampamento local restringindo os recursos das equipes.
- **Condições climáticas:** as condições climáticas podem tornar os estudos de campo arriscados impedindo sua execução.
- **Acessibilidade aos alvos:** Alvos como prédios, afloramentos e florestas, podem conter locais de difícil acesso. Um afloramento muito extenso ou alto pode conter locais onde medições se tornam impraticáveis.
- **Repetições:** Existem estudos de campo que devido a erros dos instrumentos, como a bússola, exigem que o processo seja realizado inúmeras vezes para garantir um melhor resultado.

- **Medições imprecisas:** A composição dos alvos analisados (como rochas magnéticas), podem interferir nos equipamentos utilizados em campo, podendo resultar em dados incorretos.

É possível dividir as tecnologias de *LASER scanning* em duas categorias: dinâmica e estática. Em casos de *LASER scanning* dinâmico, o aparelho é fixado em alguma plataforma móvel. Este tipo de sistema necessita de equipamentos de posicionamento (GPS) extra tornando todo sistema mais caro e complexo. Alguns exemplos possíveis são a utilização do *LASER scanner* em um automóvel ou utilizando de forma aérea em aviões e helicópteros. A categoria de *LASER scanner* estático mantém o equipamento em uma posição fixa durante todo o processo de aquisição dos dados. As grandes vantagens deste modo de trabalho são sua precisão milimétrica e densidade de dados gerados.

Diversos formatos de operação estão disponíveis para instrumentos de *LASER scanner* estáticos, variando entre três possibilidades de alcance: longo, médio e curto (VAN GENECHTEN, 2008). Inicialmente *LASER scanners* eram utilizados na indústria automotiva para mapeamentos de curto alcance facilitando o processo de prototipagem. Com a evolução tecnológica, equipamentos de médio e longo alcance foram criados, e com isso, outras áreas foram exploradas, tais como: geologia, arquitetura, engenharia civil, entre outros (Figura 6).

Figura 6: Aplicações do *LASER Scanner*



Fonte: Imagem produzida pelo autor

- **Curto Alcance:** equipamentos que tipicamente têm alcance de poucos metros com precisão submilimétrica, utilizados para modelar pequenos objetos ou artefatos em laboratório.
- **Médio Alcance:** equipamentos que utilizam o princípio de tempo de voo, com alcande entre dois e mil metros, com precisões na ordem de cinco milímetros.
- **Longo Alcance:** equipamentos com alcance de até dois quilômetros com precisão na ordem de centímetros.

LASER scanners terrestres utilizam a tecnologia LIDAR (*LIght Detection And Ranging*) para determinar a distância de um objeto distante. Tipicamente em equipamentos que utilizam esta tecnologia um pulso LASER (com comprimento de onda próximo ao infra-vermelho) é disparado contra o objeto que se deseja mapear (Figura 7). Através de um receptor é possível capturar o pulso refletido e assim obter a informação de distância do objeto.

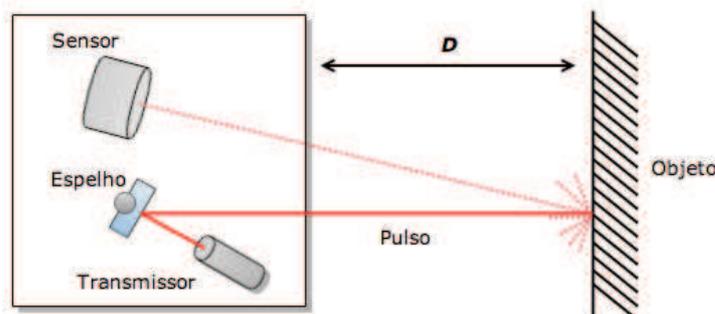
Scanners que utilizam LIDAR emitem um tipo de onda eletromagnética como luz ou raio-x. Estes equipamentos requerem um *hardware* de transmissão e recepção específicos tornando-os mais caros e complexos. Suas principais características são:

- Não requerem luz ambiente para funcionar;
- Fornecem uma densa quantidade de dados sobre o objeto mapeado (coordenadas tridimensionais) de forma automatizada;
- Eficiente aquisição dos dados, entre 1.000 e 500.000 pontos por segundo (dependendo do aparelho e configuração do mapeamento).

Um dos princípios que utilizam LIDAR para determinar a posição de um objeto é o tempo de voo utilizando pulsos *laser*. Este tipo de medição se baseia no tempo em que um pulso leva para viajar até o objeto e ser refletido de volta (*round trip*). A velocidade de obtenção dos dados utilizando este tipo de equipamento está diretamente ligada ao tempo de envio e retorno do laser.

Ao utilizar técnicas como LIDAR milhares de pontos são amostrados por segundo gerando uma densa quantidade de informações. Uma característica importante do *LASER scanner* é sua capacidade de mapear objetos tendo como resultado uma nuvem de pontos tridimensional de alta resolução (FRÖHLICH; METTENLEITER, 2004). A nuvem de pontos é a representação mais básica dos dados coletados com o *laser scanner*. Na Seção 2.2.1 mais detalhes sobre este formato serão apresentados.

Figura 7: Os sistemas que utilizam LIDAR contêm um transmissor que emite o pulso LASER. Quando o pulso colide com o objeto alvo é refletido e capturado pelo receptor do equipamento.



Fonte: Imagem produzida pelo autor

2.2.1 Nuvem de Pontos

A nuvem de pontos é o conjunto de todos os pontos tridimensionais resultantes do mapeamento com LASER scanner. A Figura 8 representa a nuvem de pontos obtida após o mapeamento de um afloramento.

Figura 8: Nuvem de pontos de um afloramento com o atributo de cor obtido pela câmera acoplada (a), e uma aproximação dos pontos (b).



(a) Nuvem de pontos



(b) Zoom nos pontos

Fonte: Imagem produzida pelo autor

Esta é a representação mais básica dos dados capturados pelo TLS. Cada ponto corresponde a uma posição (x, y, z) no espaço. Dependendo do equipamento utilizado um ou mais atributos associados aos pontos podem ser coletados tais como: intensidade do LASER e informações de cor (ALAN K. L. KWONG; WONG, 2007; VOSSELMAN et al., 2004).

O atributo de intensidade do LASER está relacionado com a potência de retorno do pulso LASER após reflexão e retorno ao sensor do scanner. O valor de intensidade depende da refletividade do objeto, distância do alvo, condições atmosféricas e possíveis obstáculos (QIHONG et al., 2012). Apesar de algumas limitações a informação de intensidade do LASER ainda é muito utilizada para estudo da litologia¹ das rochas, onde algumas informações litográficas são fortemente marcadas (HODGETTS, 2013).

O atributo de cor de cada ponto da nuvem refere-se a cor do objeto mapeado naquele ponto. Esta informação é usualmente extraída de imagens digitais obtidas pelo próprio equipamento (TLS) que tem câmera digital acoplada. A cor de cada ponto é formada por três valores representando o canal RGB (*Red, Green, Blue*). Utilizando técnicas de processamento de imagem as informações são mapeadas para cada ponto da nuvem de pontos. Com isso podendo-se recriar

¹Litologia de uma rocha é a descrição das características físicas visíveis em um afloramento, tais como cor, textura, granularidade ou composição (JACKSON; MEHL; NEUENDORF, 2005).

o objeto mapeado em um modelo digital, como ilustra a Figura 8.

A vantagem em utilizar a nuvem de pontos é que requer o menor tempo de processamento e é o tipo de dados que pode ser renderizado de forma mais rápida. Isto permite que um conjunto de dados muito maior possa ser visualizado facilmente (HODGETTS, 2013)

2.2.2 Atributos de orientação de planos

As informações obtidas através do TLS auxiliam na interpretação de estruturas geológicas (BALTSAVIAS et al., 2001; FABUEL-PEREZ; HODGETTS; REDFERN, 2009), como por exemplo estruturas planares (NAGALLI; FIORI; NAGALLI, 2011; STURZENEGGER; STEAD, 2009). Na Seção 2.1.1 foi abordado como as medidas de orientação do plano são realizadas em estudos de campo.

Após o mapeamento do afloramento todos os pontos podem ser georreferenciados por geólogos. Com isso é possível então calcular os valores de direção da camada e ângulo de mergulho do plano (ALAN K. L. KWONG; WONG, 2007). O cálculo dos valores de orientação do plano são definidos como uma relação das coordenadas tridimensionais do vetor normal do plano. Com o conhecimento prévio dos pontos que representam um plano e utilizando uma técnica de análise de dados como a PCA (Seção 2.3), pode-se extrair o vetor normal do plano.

Sendo α e β os valores de direção da camada e ângulo de mergulho do plano, respectivamente, e x , y e z as coordenadas tridimensionais do vetor normal do plano, a fórmula de cálculo dos valores de orientação do plano são dados por:

$$\alpha = \cos^{-1} \frac{|y|}{\sqrt{x^2 + y^2}} \quad (2.1)$$

$$\beta = \cos^{-1} \frac{|z|}{\sqrt{x^2 + y^2 + z^2}} \quad (2.2)$$

A alta resolução do modelo digital facilita o cálculo de informações de superfície. Estas informações são utilizadas para estudos diversos sobre a formação dos afloramentos (ALAN K. L. KWONG; WONG, 2007; VOSSSELMAN et al., 2004).

Fernández (FERNÁNDEZ, 2005) apresenta uma abordagem para encontrar o plano de melhor ajuste para um conjunto de dados selecionados. São utilizados os parâmetros de coplanaridade (se os pontos selecionados estão todos sobre uma superfície) e colinearidade (se os pontos selecionados estão alinhados em somente uma linha). Ambos parâmetros são obtidos através da análise dos autovalores descritos por Woodcock (WOODCOCK, 1977).

2.3 Análise de Componentes Principais (PCA)

Existem alguns métodos conhecidos para obtenção das informações de orientação de planos a partir de uma nuvem de pontos (NAGALLI; FIORI; NAGALLI, 2011; FENG et al., 2001;

ALAN K. L. KWONG; WONG, 2007; FERNÁNDEZ, 2005). PCA (*Principal Component Analysis* ou *Análise de Componentes Principais*) vem sendo utilizado na solução de diversos problemas (KASS; LI, 2012; CHEN; QIAN, 2011; DANDPAT; MEHER, 2013). O presente trabalho utilizará este método para detectar automaticamente os planos que melhor se aproximam aos dados de entrada.

PCA é uma técnica de análise de dados com objetivo de reduzir a dimensionalidade de um conjunto de dados com uma grande quantidade de variáveis relacionadas, preservando a variação sua variação (JACKSON, 2005). Cada variável é uma eixo ou dimensão relacionados entre si com uma determinada variância (medida de dispersão dos valores sobre a média) (DAULTREY, 1976). Estes dados são transformados de forma a representar a mesma quantidade de variação, com o mesmo número de eixos ou dimensões, com o mesmo número de variáveis, de maneira que:

- o primeiro eixo represente a maior variância possível;
- o segundo eixo represente a segunda maior variância e não seja relacionado com o primeiro eixo;
- o terceiro eixo represente a terceira maior variância e não seja relacionado com o primeiro e segundos eixos;
- e assim por diante para cada eixo ou dimensão do conjunto de dados.

Os eixos resultantes da transformação do conjunto de dados original são suas componentes principais. Por serem eixos não correlacionados tornam a análise dos dados mais simples (JOLIFE, 2002). Mesmo um conjunto de dados de muitas variáveis, grande parte da variação destes dados estarão representados em poucos grandes eixos, sendo que os eixos de menor variação usualmente são desconsiderados no processo de interpretação (DAULTREY, 1976).

Seja $X = \{P_{i,j} \in \mathbb{R}^N; i = 1, 2, \dots, N; j = 1, 2, \dots, M\}$ um conjunto de dados de tamanho M e dimensão N . Para calcular as componentes principais associadas a X , considere os elementos de X em uma matriz $[X]$ com dimensões $N \times M$ onde suas linhas contêm os pontos $P_{i,j}$ em uma ordem qualquer.

Então, calcula-se a matriz de covariância S de dimensões $N \times N$, onde N é o número de variáveis sendo analisadas, ou seja, a dimensão do conjunto de dados X . A diagonal principal da matriz conterá a variância de cada variável. A matriz simétrica S é dada pela Equação (2.3).

$$S = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1n} \\ s_{21} & s_2^2 & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_n^2 \end{bmatrix}, \quad (2.3)$$

onde, s_i^2 é a variância da i -ésima variável e s_{ij} é a covariância entre a i -ésima e a j -ésima variável.

Com a matriz de covariância S calculada, é possível obter as componentes principais de X com N variáveis (dimensões), resultando em um vetor componente (autovetor ou *eigenvector*) para cada dimensão. Todos os autovetores gerados são ortogonais entre si. A distribuição dos dados em cada componente é indicada por um coeficiente (autovalor ou *eigenvalue*) associado com o vetor. Quanto maior o coeficiente maior é a distribuição dos dados naquela direção.

Os autovalores podem ser obtidos solucionando a Equação (4.3):

$$|\mathbf{S} - \lambda\mathbf{I}| = 0, \quad (2.4)$$

onde, \mathbf{S} é a matriz de covariância, λ são os autovalores a serem calculados e \mathbf{I} é a matriz identidade. O número de valores encontrados para λ que satisfaz a equação depende da dimensão da matriz de covariância, ou seja, o número de dimensões dos dados de entrada.

Para cada autovalor encontrado calcula-se os vetores associados, como desmonstra a Equação (4.4).

$$|\mathbf{S} - \lambda_i\mathbf{I}|\mathbf{t}_i = 0, \quad i = 1, 2, \dots, N, \quad (2.5)$$

onde, \mathbf{t}_i é o autovetor N -dimensional resultante para o i -ésimo autovalor. Com os autovetores normalizados, ou seja, $|\mathbf{t}_i| = 1$, estes são ordenados em ordem decrescente baseados no autovalor associado a eles ($\lambda_1 \geq \lambda_2 \geq \dots \lambda_n$). Uma matriz TF é construída de modo que cada coluna da matriz contenha um dos vetores já ordenados, como mostra a Equação (2.6). Estes vetores definem uma nova base N -dimensional e TF é a matriz de transformação dos dados para a nova base.

$$TF = \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \dots & \vec{e}_n \end{bmatrix}, \quad (2.6)$$

onde TF é a matriz de transformação, \vec{e}_1 é o primeiro autovetor (associado ao maior autovalor), \vec{e}_2 é o segundo autovetor (associado ao segundo maior autovalor) e \vec{e}_n é o n -ésimo autovetor (associado ao menor autovalor).

Com a matriz de transformação criada torna-se simples alterar os dados de maneira a facilitar sua análise, usualmente considerando somente alguns dos componentes mais importantes dos dados. Existem diversas regras para tomar a decisão de quantos componentes são necessários para analisar os dados de entrada, preservando um certo grau de confiabilidade (JOLLIFE, 2002), tais como: o percentual cumulativo do total de variação, as grandezas de variação das componentes principais, o diagrama LEV (log-eigenvalue), entre outros.

Remover algumas variáveis da análise dos dados implica em perda de informação. Contudo, utilizando a técnica de análise de componentes principais decidir quais variáveis serão excluídas, mantendo um certo grau de confiabilidade, é simples e com baixa implicação no resultado final.

2.3.1 Planos com PCA

Este trabalho tem como objetivo calcular os melhores planos que aproximam a nuvem de pontos obtida através do LASER *scanner* terrestre. Para o método de PCA as variáveis são os eixos X, Y e Z do espaço cartesiano, onde seus valores são a posição de cada ponto. O método terá como saída três componentes principais indicando as maiores variações dos dados. As componentes principais resultantes do método de PCA não são relacionadas entre si, logo os eixos resultantes são ortogonais intersectando-se no ponto médio dos dados de entrada.

É possível definir um plano em um espaço tridimensional de diversas formas, entre elas:

- **Utilizando um ponto sobre o plano e o vetor normal deste plano:** Sendo P_0 um ponto qualquer sobre o plano e \vec{n} o vetor normal deste plano, é possível saber se um ponto P está sobre o plano se somente se o vetor entre os pontos P e P_0 é perpendicular ao vetor normal \vec{n} . Sabe-se que um vetor é perpendicular a outro se somente se o produto escalar entre os vetores é zero, logo temos a Equação (2.7).

$$\vec{n} \cdot (P - P_0) = 0 \quad (2.7)$$

- **Utilizando um ponto e dois vetores sobre o plano:** Sendo \vec{v} e \vec{z} dois vetores não paralelos sobre o plano e P_0 um ponto qualquer sobre o plano, é possível descrever parametricamente qualquer ponto P sobre o plano, de acordo com a Equação (2.8).

$$P = P_0 + a\vec{v} + b\vec{z}, \quad \forall a, b \in \mathbb{R} \quad (2.8)$$

Seja $Y = \{P_i(x_i, y_i, z_i) \in \mathbb{R}^3; i = 1, 2, \dots, N\}$ um conjunto tridimensional de pontos, podemos aplicar o método de PCA descrito na Seção 2.3, onde a dimensão do conjunto Y de entrada é 3. A propriedade importante obtida desta análise é que os autovetores resultantes do processo de análise de componentes principais não são correlacionados em relação aos dados de Y . Os autovalores associados representam a medida de desvio padrão dos dados do conjunto Y em relação aos autovetores.

Desta forma, o autovetor associado ao autovalor λ_1 juntamente com o ponto médio \bar{P} representam a linha de melhor ajuste aos dados do conjunto Y . Quando são utilizados os autovetores associados aos dois maiores autovalores (λ_1, λ_2) e o ponto médio \bar{P} o plano de melhor ajuste aos dados do conjunto Y é obtido. Em ambos os casos, o melhor ajuste se refere ao ajuste ortogonal dos dados, i.e., minimizando a soma dos quadrados das distâncias dos pontos com a

linha ou plano. O menor autovalor, aquele associado ao menor autovalor (λ_3), mede a dispersão dos pontos de Y em relação ao plano de melhor ajuste, e pode ser utilizado como uma medida de qualidade do plano.

Sendo assim, utilizando o método de análise de componentes principais e selecionando seus vetores temos:

- O plano tem as menores distâncias ortogonais para cada ponto de entrada;
- Os vetores sobre o plano não são paralelos, logo podemos definir parametricamente o plano;
- O ponto de intersecção dos vetores é a média dos dados de entrada (utilizado no método de PCA);
- O terceiro vetor resultante o método \vec{e}_3 é a normal do plano, pois por definição da PCA, este é perpendicular aos outros vetores.

3 TRABALHOS RELACIONADOS

Modelos digitais de afloramentos são usualmente compostos de uma grande densidade de dados para serem interpretadas. A capacidade de analisar estes dados de forma ágil e precisa com facilidade é importante para criação de grandes bases de dados geológicas e geoestatísticas.

Este capítulo apresentará de forma breve alguns trabalhos relacionados a detecção de planos e análise de afloramentos.

3.1 Detecção de planos

Usualmente métodos manuais de delimitação de planos, fraturas e outras estruturas planares importantes, são aplicadas no modelo digital do afloramento (HODGETTS, 2013). Uma interpretação mais avançada requer as informações de orientação dos planos, tais as descritas na Seção 2.1.1. Os valores de orientação de estruturas planares em afloramentos são usualmente medidas utilizando um compasso e clinômetro. As desvantagens em utilizar este método é que ele pode levar um longo período de tempo quando diversas estruturas são medidas, devido ao tamanho do afloramento. Também pode ser impossível de realizar a medida quando a face da rocha está fisicamente inacessível (FENG et al., 2001).

É possível extrair tais informações do modelo digital de afloramento utilizando técnicas matemáticas sobre os dados de entrada e obtendo informações sobre os planos, como já apresentado na Seção 2.3.1.

Técnicas manuais de interpretação de dados, mais precisamente de orientação de estruturas planares, requerem que o geólogo selecione algumas informações do modelo digital. Após a seleção prévia dos dados de entrada a técnica desejada é aplicada e os resultados obtidos. Algumas destas técnicas são:

- **Três pontos:** selecionando três pontos não colineares na nuvem de pontos, no local onde deseja-se obter a informação de orientação da estrutura planar, é possível calcular a normal do plano e obter as informações de orientação. A seleção de três pontos faz com que exista somente um plano no espaço tridimensional onde estes três pontos pertençam. Sendo P_0 , P_1 e P_2 os pontos selecionados e \vec{v}_1 e \vec{v}_2 os vetores P_0P_1 e P_0P_2 , respectivamente. O vetor normal \vec{n} é o resultado do produto vetorial entre \vec{v}_1 e \vec{v}_2 .

$$\vec{n} = \vec{v}_1 \times \vec{v}_2 \quad (3.1)$$

- *Tensor method:* este método tem como entrada um conjunto de dados para realizar o cálculo do plano de melhor ajuste aos dados. Woodcock (WOODCOCK, 1977) descreve o método e o que os valores encontrados de grau de ajuste e confiabilidade representam em relação aos dados de entrada, ou seja, descrevem a forma dos dados de entrada. Fernández (FERNÁNDEZ, 2005) realiza um estudo avaliando os possíveis resultados obtidos

com o *tensor method* com o objetivo de identificar formas planares e que possam ser utilizados em modelos digitais de afloramentos. São definidos então os valores indicados de confiabilidade do método proposto por Woodcock (WOODCOCK, 1977) para aplicação em modelos digitais de afloramento.

- **Regressão planar:** Feng (FENG et al., 2001) utilizou a técnica dos mínimos quadrados com regressões lineares multivariadas para detectar o plano de melhor ajuste em uma seleção de pontos. A Equação (3.2) expressa o plano de melhor ajuste.

$$z = b_0 + b_1x + b_2y \quad (3.2)$$

Utilizando o método dos mínimos quadrados, assume-se que o plano de melhor ajuste é gerado quando a menor soma dos quadrados é encontrada (equação 3.3).

$$\min \epsilon^2 = \sum [z_i - (b_0 + b_1x_i + b_2y_i)]^2 \quad (3.3)$$

Souza (SOUZA, 2012) realizou um estudo comparando os métodos de três pontos, regressão planar e *tensor* entre si, e com os valores obtidos em campo com bússola e clinômetro. Para o método *tensor* os valores sugeridos por Fernandez (FERNÁNDEZ, 2005) para melhor ajuste do plano são utilizados e a partir dos resultados do método os parâmetros de orientação do plano são calculados. Na análise dos resultados foi verificado que todos os métodos foram eficazes no cálculo dos valores de orientação dos planos. Em comparação com os valores de campo (medidos com bússola e clinômetro) ocorreram diferenças que se mantiveram entre $\pm 3^\circ$.

3.2 Análises geológicas automáticas

A interpretação de dados demandam uma quantidade de tempo proporcional ao volume de dados. Existem diversos estudos para o desenvolvimento de análises automatizadas com o objetivo de reduzir o tempo utilizado na coleta de informações geostatísticas.

Nas ciências geológicas, as análises realizadas sobre os dados requerem conhecimentos especializados adquiridos com diversos anos de aprendizagem e experiência. Para o desenvolvimento de técnicas de análise mais automatizadas vem sendo desenvolvido métodos utilizando inteligência artificial (IA) (HODGETTS, 2013).

Inteligência *swarm* é uma técnica de IA que se baseia no comportamento coletivo detectado em grupos de insetos, mais precisamente formigas, onde simples ações individuais de cada membro resultam na solução de problemas complexos (BONABEAU; DORIGO; THERAULA, 1999). Este tipo de técnica vem sendo aplicado em interpretações para mapeamento de falhas em volumes sísmicos 3D e fraturas em dados obtidos com o *laser scanner* (PEDERSEN et al., 2005).

Redes neurais artificiais são uma linha de inteligência artificial que se propõe a resolver problemas de maneira similar a sistemas biológicos como o cérebro. Estes modelos tem a capacidade de aprendizado, e se destacam em reconhecimento de padrões e classificação (RUSSELL et al., 1995). Este tipo de técnica tem sido integrada com os dados obtidos do *laser scanner* e imagens multiespectrais para caracterização de solos pantanosos (MORRIS et al., 2005).

O desenvolvimento de técnicas de reconhecimento de padrões geológicos traz grandes avanços para a interpretação de dados digitais. Tais métodos utilizam técnicas de aprendizagem sobre um determinado conjunto de informações. Com isso resultados diferentes são obtidos para um mesmo conjunto de informações. O reconhecimento de estruturas planares pode ser realizado utilizando técnicas matemáticas mais objetivas. Com isso os resultados serão consistentes para um mesmo conjunto de informações. Análise de Componentes Principais (Seção 2.3) pode ser utilizada para detecção destas estruturas de maneira eficiente.

4 DETECÇÃO AUTOMÁTICA DE PLANOS EM AFLORAMENTOS DIGITAIS

O objetivo principal deste trabalho é a descoberta automática de planos em uma nuvem de pontos de afloramentos utilizando a técnica de PCA. Contudo, esta técnica pode ser aplicada a uma nuvem de pontos qualquer. O processo se baseia em encontrar um grupo de pontos os quais sejam uma boa aproximação a um plano geométrico. Os resultados obtidos pela PCA, vistos na Seção 2.3, são vetores e coeficientes que indicam a distribuição dos pontos no espaço. Com estes dados, podemos obter a definição do melhor plano que aproxima aqueles pontos e o quão adequado os pontos estão de um plano real. Este capítulo apresenta em maiores detalhes o processo de cálculo dos planos em uma nuvem de pontos utilizando a técnica de PCA e o método desenvolvido para detecção automática dos diversos planos em um modelo digital de afloramentos.

4.1 Calculando o plano de melhor ajuste

O método de análise de componentes principais tem como resultado autovetores e autovalores que indicam a distribuição de um conjunto de dados, conforme apresentado na Seção 2.3. Ao aplicarmos esta técnica nos dados do modelo digital de afloramentos, teremos que cada coordenada da nuvem de pontos serão as variáveis (dimensões) analisadas. Como já apresentado na Seção 2.3.1, os vetores resultantes do método de PCA são ortogonais entre si intersectando-se no ponto médio dos dados de entrada. Os vetores que indicam a duas maiores direções de dispersão dos dados serão utilizados para definir o plano que melhor aproxima a nuvem de pontos. Nesta seção serão apresentados os detalhes de cálculo realizados neste processo.

A nuvem de pontos é a representação básica do modelo digital de afloramentos, onde cada ponto é representado por uma posição (x, y, z) no espaço (Seção 2.2.1). As três coordenadas dos pontos de entrada serão as variáveis de entrada do método de PCA, e o resultado indicará três vetores de dispersão destes pontos.

Seja $X = \{P_i(x_i, y_i, z_i) \in \mathbb{R}^3; i = 1, 2, \dots, N\}$ um conjunto de pontos no espaço tridimensional. Para calcular os componentes principais associados a X , uma matriz $[X]_{N \times 3}$ é criada, onde as linhas são os pontos P_i , em uma ordem qualquer. Desta forma, as três colunas de $[X]$ são os vetores \vec{x} , \vec{y} e \vec{z} , com os valores das variáveis x , y e z , respectivamente.

Como introduzido na Seção 2.3 a matrix de covariância $S_{3 \times 3}$ é definida conforme a Equação (4.1).

$$\mathbf{S} = \begin{bmatrix} s_x^2 & s_{xy} & s_{xz} \\ s_{yx} & s_y^2 & s_{yz} \\ s_{zx} & s_{zy} & s_z^2 \end{bmatrix}, \quad (4.1)$$

onde, s_i^2 é a variância da **i-ésima** variável e s_{ij} é a covariância entre a **i-ésima** e **j-ésima** variável,

definido como:

$$s_{ab} = \frac{N \sum_{n=1}^N a_n b_n - \sum_{k=1}^N a_k \sum_{k=1}^N b_k}{N(N-1)}, \quad (4.2)$$

onde, a e b representam x ou y .

Após o cálculo da matriz S é possível calcular os autovalores solucionando a Equação (4.3).

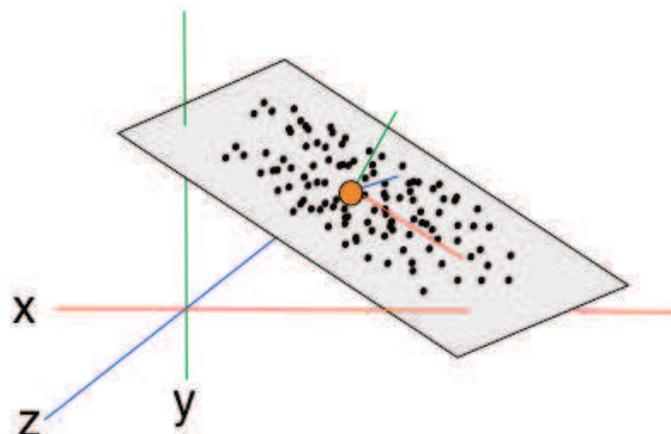
$$|\mathbf{S} - \lambda \mathbf{I}| = 0, \quad (4.3)$$

onde λ são os autovalores a serem calculados e \mathbf{I} é a matriz identidade. Solucionando a Equação (4.4) obtém-se os autovetores associados aos autovalores λ_i .

$$[\mathbf{S} - \lambda_i \mathbf{I}] \mathbf{t}_i = \mathbf{0}, \quad i = 1, 2, 3, \quad (4.4)$$

onde \mathbf{t}_i é o autovetor resultante para o i -ésimo autovalor. Os autovetores são normalizados, ou seja, $|\mathbf{t}_i| = 1$, e os vetores \mathbf{t}_i são ordenados, em ordem decrescente, baseados nos autovalores associados a eles ($\lambda_1 \geq \lambda_2 \geq \lambda_3$). Como apresentado na Seção 2.3, os autovetores resultantes são ortogonais, formando uma base em \mathbb{R}^3 . Os dois maiores autovetores, ou seja, aqueles associados aos dois maiores autovalores, descrevem as direções de maior dispersão dos dados, portanto definindo um plano. O terceiro autovetor, associado ao menor autovalor, é a normal deste plano.

Figura 9: Vetores obtidos da técnica de PCA formando um plano com sua normal.

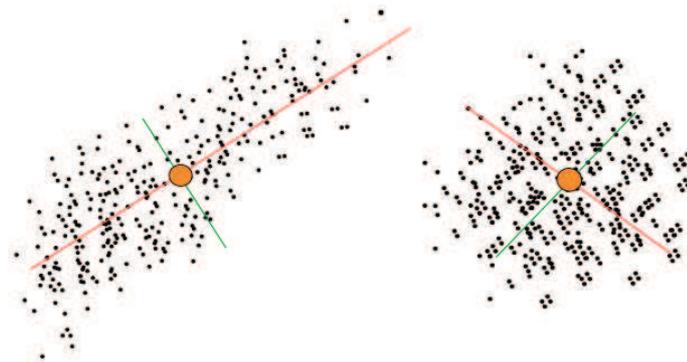


Fonte: Imagem produzida pelo autor

Aplicando o método de PCA aos dados de entrada, são obtidos os vetores que em conjunto com o ponto médio dos dados representam um plano geométrico. Mesmo sabendo que a técnica de PCA produz a melhor aproximação geométrica possível para os dados, ou seja, um plano que

está localizado com a menor distância ortogonal de todos os dados, não é garantido que os dados de entrada realmente representam um plano (Figura 11).

Figura 10: Representação de uma visão lateral da nuvem de pontos, onde a linha verde representa a normal do plano e a linha vermelha a direção de maior dispersão dos dados. A esquerda os pontos tem uma tendência a representarem um plano, contudo, a direita os dados estão igualmente distribuídos no espaço indicando que não representam um plano.



Fonte: Imagem produzida pelo autor

Uma boa aproximação de um plano é quando a variação dos dados em relação ao vetor normal do plano é a menor possível. Isto indica que os pontos estão alinhados, formando uma superfície plana. Woodcock (WOODCOCK, 1977) apresenta uma análise dos autovalores para determinar a forma associada a uma nuvem de pontos. Duas informações, coplanaridade e colinearidade, são obtidas para determinar a forma. Com estas informações é possível determinar o quão ajustados os dados de entrada estão ao plano encontrado pelo método de PCA.

A coplanaridade ou grau de ajuste é representado por m , e pode ser obtido pela proporção entre o maior autovalor λ_1 e o menor autovalor λ_3 , como descreve a Equação (4.5).

$$m = \ln\left(\frac{\lambda_1}{\lambda_3}\right) \quad (4.5)$$

Quanto maior o valor de m , menor é a distância entre os pontos de entrada e o plano estimado. Contudo, um bom ajuste dos pontos não é o suficiente para determinar a qualidade do plano. Os dados de entrada mesmo estando bem próximos ao plano aproximado, podem não estar bem distribuídos sobre a superfície. Com isso avaliar a colinearidade dos dados é necessário.

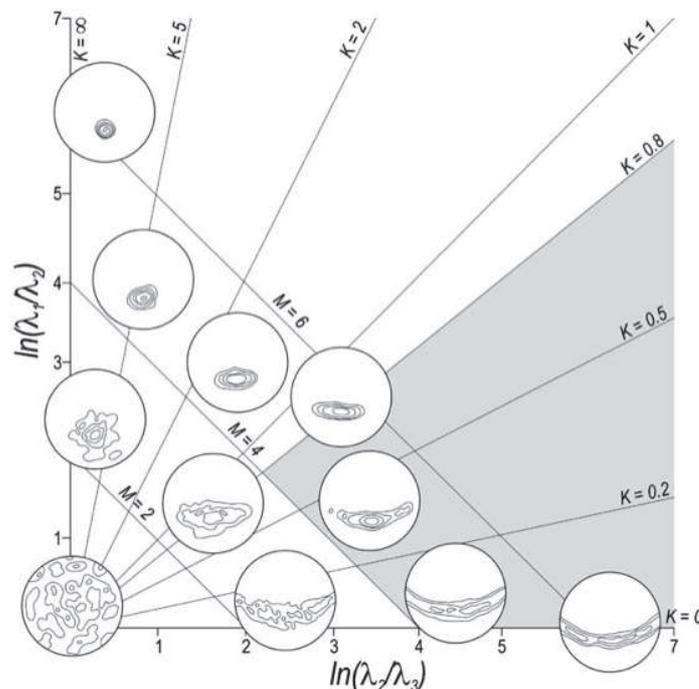
A colinearidade ou grau de confiança é representado por k , e pode ser obtido com a relação entre os três autovalores, como descreve a Equação (4.6).

$$k = \frac{\ln\left(\frac{\lambda_1}{\lambda_2}\right)}{\ln\left(\frac{\lambda_2}{\lambda_3}\right)} \quad (4.6)$$

Quanto menor o valor de k mais homogeneamente distribuídos estão os pontos sobre a superfície do plano. Isto indica que a distribuição dos dados não se aproxima de uma reta no espaço. O desejado é que a dispersão dos dados seja similar na direção dos vetores que representam o plano.

Estes parâmetros (WOODCOCK, 1977) podem resultar em diversos valores, Fernández (FERNÁNDEZ, 2005) descreve a relação entre os valores de m e k (Figura 11) e a distribuição dos pontos no espaço. Dois valores de referência são apresentados para o grau de ajuste e confiabilidade. Fernández destacou que, para o melhor ajuste dos dados ao plano, m deve ser maior ou igual que quatro, indicando uma variação de até seis graus acima da superfície do plano. Para uma melhor confiança na distribuição dos dados foi identificado que k deve ter valores menores ou iguais que 0,8 indicando uma distribuição sobre a superfície do plano maior que trinta graus em relação ao centro dos dados.

Figura 11: Valores de ajuste e confiabilidade dos planos de acordo com a proporção dos autovalores obtidos.



Fonte: Imagem retirada de (FERNÁNDEZ, 2005)

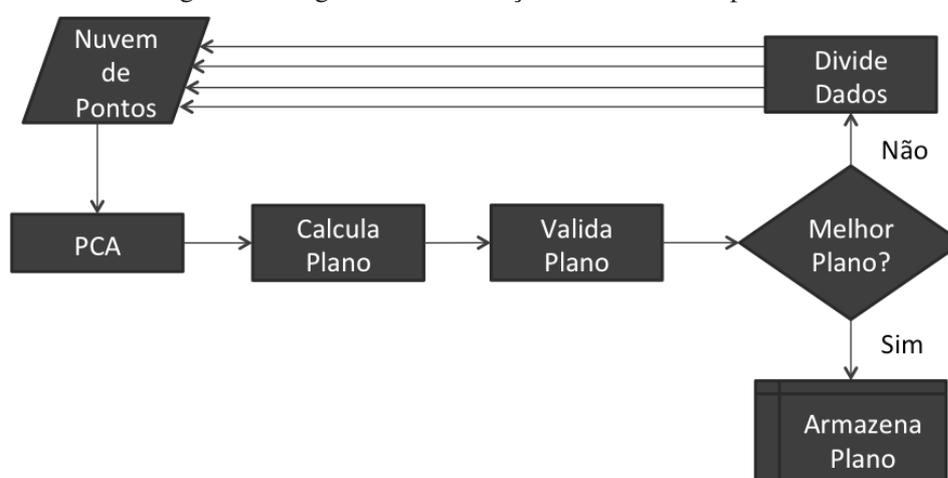
4.2 Detecção automática de planos

O principal objetivo deste trabalho é detectar, de forma automática, planos em modelos digitais de afloramentos. Na Seção 4.1 foi apresentado como a técnica de análise de componentes principais é utilizada para calcular a melhor aproximação de um plano para uma nuvem de pontos selecionada. Para atingir o objetivo deste trabalho iremos utilizar os valores de grau

de ajuste m e confiabilidade k do plano para tomar a decisão de divisão da nuvem de pontos (Figura 12).

O processo para detecção de planos tem como entrada toda a nuvem de pontos obtidas através do mapeamento de um afloramento. Utilizando o método de PCA, o melhor plano que aproxima tais dados será encontrado, conforme descrito na Seção 4.1. A partir da informação do plano são calculados os valores de grau de ajuste e confiabilidade, e um teste é realizado validando se os valores encontrados estão dentro do mínimo esperado. Enquanto o teste for negativo os dados de entrada são divididos em quatro partes e todo o processo é repetido para cada parte até que um plano seja validado (Figura 12).

Figura 12: Algoritmo de detecção automática de planos.



Fonte: Imagem produzida pelo autor

4.2.1 Divisão dos dados de entrada

O processo de aproximação do plano é realizado utilizando a técnica de PCA. É necessário que, após a validação negativa do plano, os dados de entrada sejam divididos. É interessante que a divisão seja realizada de forma que os valores de dispersão dos pontos fiquem cada vez menores. Ou seja, a divisão é baseada nos dois vetores que compõem o plano, pois estes indicam as duas maiores direções de dispersão dos dados.

Como dois vetores serão utilizados para dividir os pontos, o resultado final serão quatro grupos de pontos (nuvens de pontos) que serão novamente entrada para a aproximação do plano utilizando PCA. Abaixo são listadas algumas definições utilizadas para a divisão dos dados:

- Seja $E = \{P_i(x_i, y_i, z_i) \in \mathbb{R}^3; i = 1, 2, \dots, N\}$ os pontos de entrada do processo;
- Sejam Q_0, Q_1, Q_2 e Q_3 os conjuntos de dados resultantes da divisão dos dados de E ;
- Sejam \vec{e}_0 e \vec{e}_1 os vetores que indicam as duas direções de maior dispersão dos dados de E ;

- Seja P_m o ponto médio dos dados de E .

Para determinar qual grupo pertence um ponto P_i qualquer, primeiramente é necessário transladar a nuvem de pontos para o centróide, fazendo com que ele seja transformado em um ponto próximo aos vetores \vec{e}_0 e \vec{e}_1 normalizados. Logo temos que:

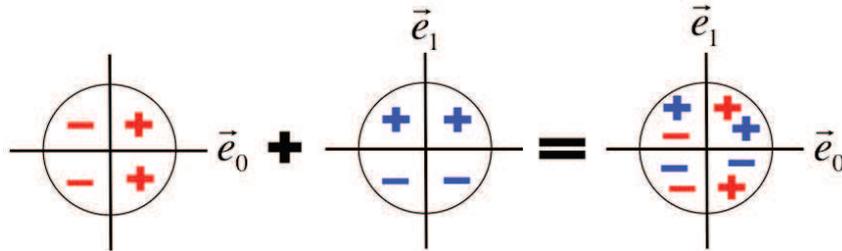
$$P' = (P_i - P_m) \quad (4.7)$$

Com o ponto transformado, a decisão de qual grupo, Q_0, Q_1, Q_2 ou Q_3 , o ponto pertence é tomada calculando o produto escalar do novo ponto P' com cada um dos vetores \vec{e}_0 e \vec{e}_1 . Sabemos que o resultado do produto escalar é o cosseno do ângulo entre os dois vetores, logo:

$$P' \cdot \vec{e}_i = |P'| |\vec{e}_i| \cos \theta \quad (4.8)$$

Os quadrantes do cosseno no círculo trigonométrico são divididos de acordo com o sinal do valor resultante. Utilizando esta informação para o resultado do produto escalar entre o ponto P' e os vetores \vec{e}_1 e \vec{e}_2 teremos como resultado um círculo trigonométrico como ilustrado na Figura 13.

Figura 13: A divisão dos pontos será referente ao enquadramento dos valores dos produtos escalares com ambos os vetores.



Fonte: Imagem produzida pelo autor

Com estas informações foram estabelecidas regras para definir a qual grupo o ponto pertence, sendo elas:

$$\begin{cases} P' \in Q_0 & \mathbf{SE} (P' \cdot \vec{e}_0) \geq 0 \mathbf{E} (P' \cdot \vec{e}_1) \geq 0 \\ P' \in Q_1 & \mathbf{SE} (P' \cdot \vec{e}_0) \geq 0 \mathbf{E} (P' \cdot \vec{e}_1) < 0 \\ P' \in Q_2 & \mathbf{SE} (P' \cdot \vec{e}_0) < 0 \mathbf{E} (P' \cdot \vec{e}_1) \geq 0 \\ P' \in Q_3 & \mathbf{SE} (P' \cdot \vec{e}_0) < 0 \mathbf{E} (P' \cdot \vec{e}_1) < 0 \end{cases}$$

Após todos os pontos serem alocados a seus respectivos grupos o processo descrito na Seção 4.2 é repetido em todos os grupos até que uma boa aproximação ao plano seja encontrada. Com isso a cada iteração os planos serão menores.

Durante o processo de divisão dos dados não há nenhum conhecimento prévio da organização da nuvem de pontos e nem de grupos de pontos que podem ser mais prováveis de formarem

um plano. Desta forma quando um grupo de pontos é dividido é possível que neste conjunto de dados os pontos que definem um plano sejam divididos. Isso ocasionará que um grande plano presente na nuvem de entrada seja dividido N vezes, e com isso N planos podem ser detectados. Medir mais de um plano sobre um grande conjunto de dados, que representa um grande plano, trará o benefício no detalhamento dos dados de direção da camada e ângulo de mergulho. Com isso poderão ser detectadas algumas irregularidades na estrutura geral que a nuvem representa.

4.3 Implementação

Foi construída uma aplicação para a visualização dos dados obtidos a partir do *laser scanner*. O dado padrão de entrada é a representação básica do modelo digital de afloramento, a nuvem de pontos.

Após o processo de captura, um arquivo de dados de formato proprietário é armazenado em meio eletrônico como unidades externas de armazenamento (disco rígido portátil, *pendrive*, etc.). Para manipulação destes dados, a empresa *Optech* disponibiliza uma aplicação para conversão de dados chamada *Parser*. Entre os diversos formatos de saída (pif, raw, ptx, 3dv, bwp, xyz, s3d, ptc, blf, iva, ixf), o utilizado neste trabalho é o xyz. Este formato é um arquivo texto com as coordenadas (x, y, z) de cada ponto e a informação de cor do *laser*. Sua estrutura é simples e contém as informações necessárias para visualização e estudo do modelo digital.

A implementação foi realizada na linguagem C++ utilizando a biblioteca Qt (NOKIA, 2009) para interface de usuário, a biblioteca *Open Scene Graph* (OSG) versão 3.1.3 (WANG; QIAN, 2010) para manipulação e visualização dos objetos 3D e a biblioteca Eigen (EIGEN 3.2.0 DOCUMENTATION, 2012) para os cálculos de álgebra linear.

Qt é uma biblioteca multiplataforma de UI (*User Interface*). Desenvolvida na linguagem C++, ela complementa a biblioteca de classes da linguagem incluindo ferramentas que auxiliam na rápida criação de aplicativos.

OSG é uma API de código aberto que implementa uma grande quantidade de funcionalidades, além das já esperadas, de um grafo de cena. É distribuída sobre a licença LGPL, e foi desenvolvida para ser escalável, portátil e multiplataforma. Foi desenvolvida em C++ utilizando *design patterns* facilitando sua integração em diversas aplicações e ambientes.

Eigen é uma biblioteca criada para solucionar problemas de álgebra linear tais como: matrizes, vetores e algoritmos de soluções numéricas. Desenvolvida em C++ suporta estruturas de dados de diversos tamanhos e contém algoritmos confiáveis para decomposição de matrizes e outras operações.

4.3.1 Detalhes do aplicativo

O cálculo de aproximação dos planos foi desenvolvido de acordo com as definições descritas na seção Seção 4.2. Contudo durante alguns testes foi verificado que algumas condições extras

de parada deveriam ser acrescentadas.

Como explicado, os dados são divididos até que uma boa aproximação de plano seja encontrada, contudo nem sempre serão encontrados planos que satisfaçam os parâmetros de ajuste e confiabilidade. Desta forma, assim como qualquer algoritmo recursivo, foi aplicado uma condição de parada η baseada no número de vezes que a entrada foi dividida. Como os dados são divididos em quatro partes, pode-se pensar que esta divisão forma uma *quad-tree* e que η é a altura máxima desta árvore.

Durante testes executados foi identificado a necessidade de limitar o tamanho mínimo dos dados de entrada para o método de PCA. Foi definido então o parâmetro δ , que valida o tamanho mínimo de uma divisão de um grupo.

A visualização dos dados é realizada com o auxílio da biblioteca de grafo de cena OSG. Um grafo de cena é uma estrutura de dados genérica utilizada para organizar logicamente uma cena. É composta de nodos conectados entre si formando um grafo ou árvore. Na OSG tais nodos são denominados de objetos OSG ou *OSG nodes*.

A API OSG contém a implementação para leitura de diversos formatos de arquivo (*obj* por exemplo). Ela permite que novos formatos sejam adicionados com a criação de *plugins*. Foi criado o *plugin* que realiza a leitura dos dados e construção do grafo de cena para o formato XYZ.

Toda carga do arquivo XYZ é realizada em um grande bloco para a memória minimizando o acesso a disco. Após a leitura em memória os dados são então transformados para objetos OSG e organizados em um grafo de cena.

Além dos objetos OSG com os dados da nuvem de pontos, foram incluídos outros objetos OSG para visualização dos pontos que compõem um plano, um objeto para armazenar o marcador de cada plano e um objeto HUD para visualização dos atributos do plano (ângulo de mergulho e direção da camada). A Figura 14 ilustra como o grafo de cena foi organizado.

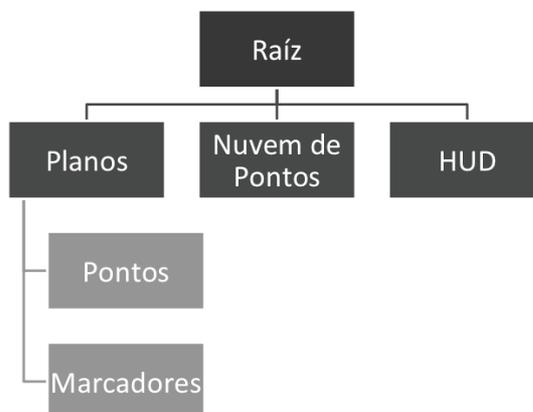
Os dados para visualização dos planos são: os pontos que deram origem ao plano e o marcador central do plano. Para auxiliar na detecção de similaridades as cores atribuídas a cada plano e seus pontos são baseadas nos valores dos atributos de ângulo de mergulho e direção da camada de cada plano. Para criar uma variação suave nas cores dos planos em relação aos valores com maior importância, uma função **sigmoid** parametrizada foi utilizada (Seção 4.3.2).

4.3.2 Função de cor dos planos

A visualização dos planos se dá através de um marcador na nuvem de pontos, este marcador fica localizado no ponto médio dos pontos que originaram o plano. Também são mantidos de forma destacada quais pontos contribuíram para o cálculo do plano. Para melhorar a visualização e identificação dos planos, a cor de cada objeto tridimensional foi parametrizada em uma função de acordo com os atributos de orientação de cada plano.

As cores de cada plano são baseadas nos valores de ângulo de mergulho e direção da ca-

Figura 14: Organização do grafo de cena criado para contemplar a visualização dos dados. HUD contém o objeto OSG para visualização de informações sobre o plano. O objeto nuvem de pontos contém os pontos da nuvem de pontos. O objeto pontos contém os pontos que foram selecionados e que fazem parte de algum plano sobre o modelo. Marcadores são os objetos com marcadores no centro de cada plano.



Fonte: Imagem produzida pelo autor

mada do plano (Seção 2.1.1). Isso auxilia o reconhecimento de padrões globais e locais no afloramento. As cores são compostas por três componentes: vermelho, verde e azul (RGB). A combinação de diferentes intensidades para cada componente produz uma cor específica. Para o componente vermelho foi atribuído um valor baixo fixo, pois ele não irá variar. Para o componente verde será atribuído uma intensidade de cor correspondente ao valor da direção da camada do plano, o componente azul receberá uma intensidade de cor correspondente ao valor de ângulo de mergulho do plano.

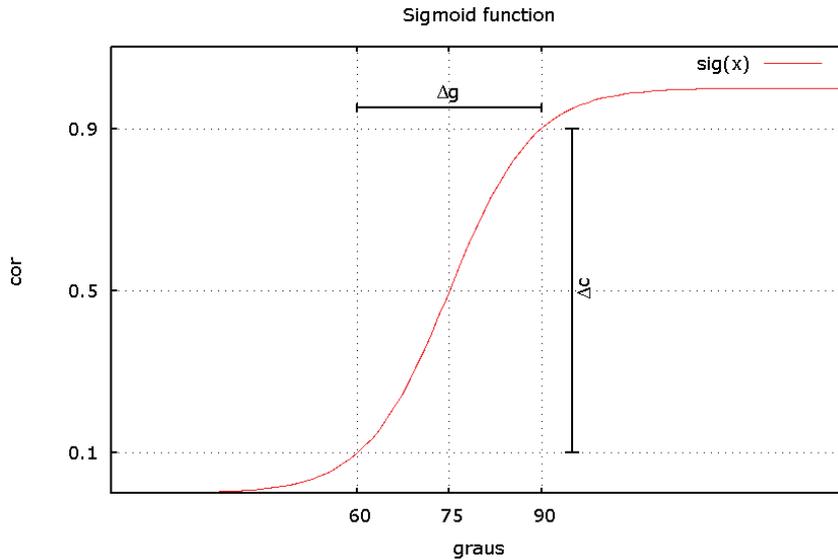
Devido a não linearidade dos valores de orientação dos planos, ou seja, em grande parte, os dados estão agrupados em um determinado intervalo de valores, foi aplicada uma função sigmoide para os componentes de cor do plano. Com esta função é possível parametrizar o intervalo de valores resultantes, tal que, a maior variação de cores seja correspondente a variação dos valores de atributo do plano. Atribuindo um determinado intervalo de valores à função, é possível diferenciar, através de cores, planos onde a inclinação esteja no intervalo de $[35^\circ, 40^\circ]$, por exemplo.

O modelo digital de afloramentos utilizado neste trabalho para fins de teste contém, em grande parte, planos que estão no intervalo de $[70^\circ, 90^\circ]$. Desta forma, a cor dos planos será mais azulada quanto maior for o valor da direção da camada do plano, e mais esverdeada quanto maior for o valor do ângulo de mergulho do plano. A Figura 15 apresenta a curva das cores do plano.

A função sigmoide é descrita pela Equação (4.9), onde x é um dos valores de orientação do plano (ângulo de mergulho ou direção da camada), ω e γ são o ponto de inflexão da curva e a inclinação da curva, respectivamente.

$$f(x) = (1 + e^{\frac{-x+\omega}{\gamma}})^{-1} \quad (4.9)$$

Figura 15: Função Sigmoid com os parâmetros de intervalo de cor Δc e parâmetros de intervalo dos valores de atributo do plano Δg .



Fonte: Imagem produzida pelo autor

A flexibilidade dada pela função sigmoide se dá devido aos seus parâmetros ω e γ . Eles permitem realizar a relação entre o intervalo de cores que deseja-se definir com o intervalo de valores para um dos atributos do plano. O objetivo é que exista a maior variação de cores possível para um determinado intervalo de valores interessantes para o estudo.

Quando o valor de ω é inserido, o que ocorre é a definição do ponto central da curva, ou seja, o ponto de inflexão da curva. Para isso simplesmente calculamos o ponto médio do intervalo de valores predeterminado de um dos atributos do plano (Equação (4.10)).

$$\omega = \frac{g_{min} + g_{max}}{2}, \quad (4.10)$$

Onde g_{min} e g_{max} são os limites superior e inferior definidos para os valores de atributos do plano.

O valor de γ define a inclinação da curva, que indica a real variação entre as cores e os valores de ângulo de mergulho e direção da camada do plano. Utilizando a função da Equação (4.9), temos que:

$$\begin{aligned} f'(x)|_{x=0} &= -(1 + e^{\frac{-x}{\gamma}})^{-2} \cdot \frac{-1}{\gamma} \\ &= \frac{1}{\gamma} \cdot \frac{1^2}{2} \\ &= \frac{1}{4\gamma} \end{aligned}$$

logo a tangente da curva é:

$$tg\theta = \frac{1}{4\gamma} \quad (4.11)$$

Desta forma é possível calcular γ utilizando os limites superiores definidos para as cores e para os valores do plano, como apresenta a Equação (4.12).

$$\gamma = \frac{1}{4} \cdot \frac{g_{max} - g_{min}}{c_{max} - c_{min}}, \quad (4.12)$$

Onde c_{max} e c_{min} são os limites superiores e inferiores definidos para variação das cores. A função é aplicada para os valores de ângulo de mergulho e direção da camada, que representam as componentes azul e verde, respectivamente.

5 RESULTADOS

Para avaliação do desempenho do algoritmo desenvolvido, testes foram realizados em superfícies sintéticas (Seção 5.1) e dados de modelos digitais de afloramentos (Seção 5.2). Os dados criados para simular uma superfície tem como objetivo validar a precisão obtida variando o parâmetro de grau de ajuste dos planos. Os testes realizados sobre o modelo digital tem o intuito de validar a detecção dos planos e precisão dos valores de ângulo de mergulho e direção da camada destes planos. Nas tabelas comparativas da Seção 5.2, a direção da camada será referenciada pela sigla *DD* e o ângulo de mergulho pela sigla *DA*.

5.1 Superfície sintética

Alguns testes foram realizados utilizando dados de superfície gerados a partir de uma função. O objetivo destes testes era avaliar a precisão e as mudanças que ocorriam no algoritmo quando o parâmetro de ajuste dos pontos ao plano m era alterado.

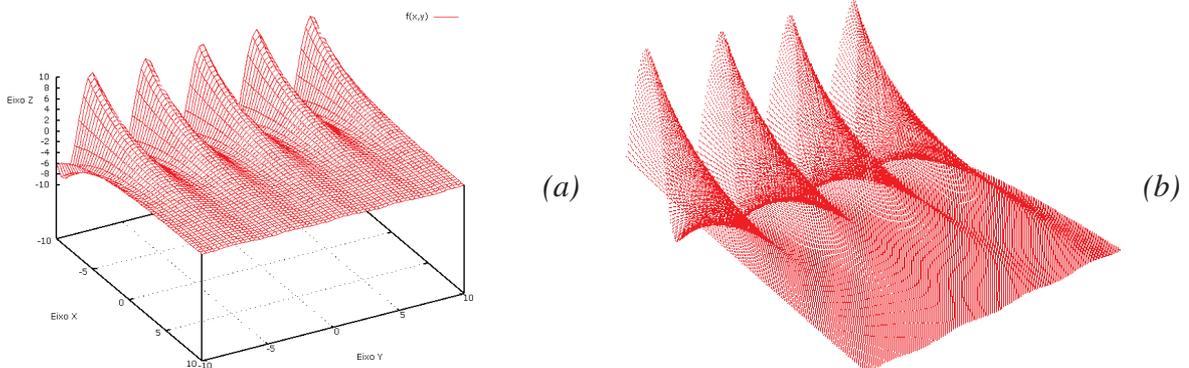
Uma aplicação foi criada para gerar uma nuvem de pontos com os dados de superfície baseados em uma função f . O arquivo de saída é gerado no formato XYZ descrito anteriormente, mas com uma única cor para cada ponto (foi selecionada a cor vermelha para melhor visualização).

A função f descreve a superfície como uma variação nos valores no eixo z baseando-se nos valores dos eixos x e y , utilizando a seguinte equação:

$$z = \rho^x \cdot \sin(\psi \cdot y) \quad (5.1)$$

Esta função foi escolhida devido a variação da posição dos pontos, que tendem gradativamente a um plano. Com estes dados será possível identificar a precisão dada pelo parâmetro m . A Figura 17 ilustra a forma da função descrita e os dados renderizados, para $\rho = 0,8$ e $\psi = 1,5$.

Figura 16: *Plot* da função f (a) e renderização do arquivo com a nuvem de pontos gerada (b).

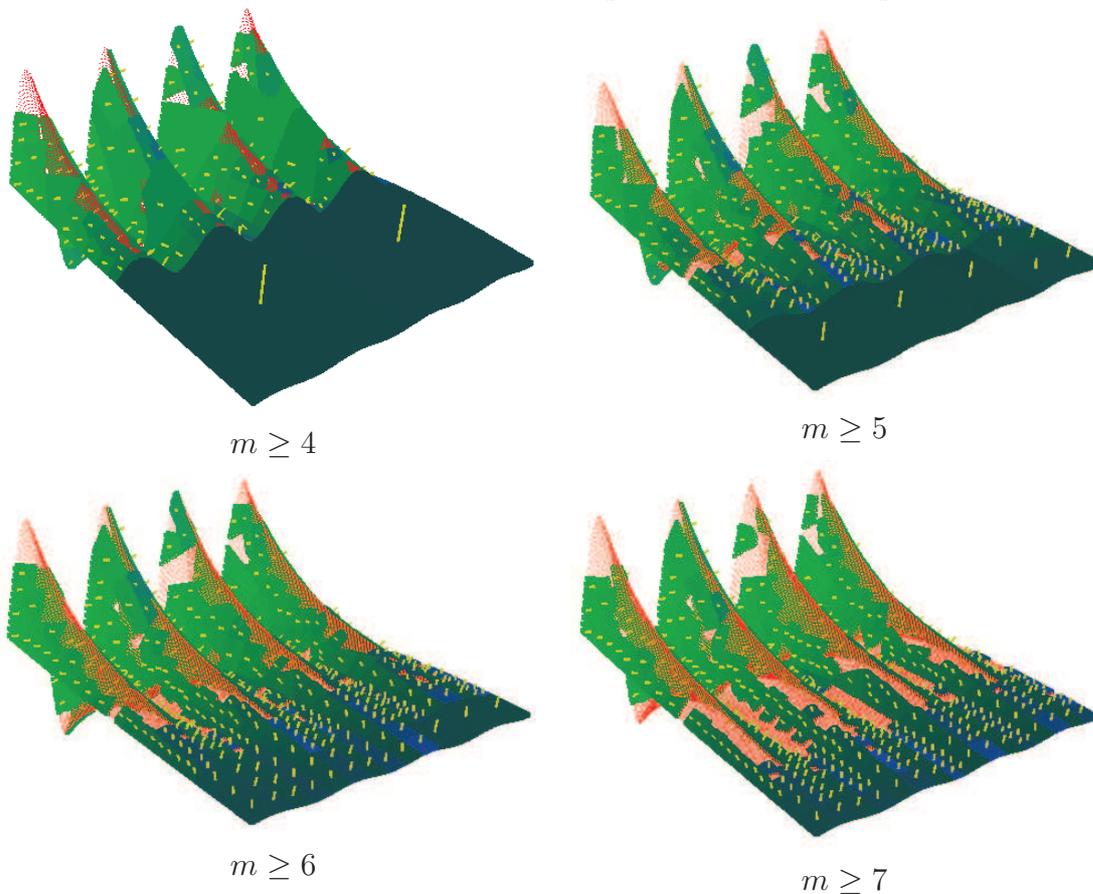


Fonte: Imagem produzida pelo autor

Com a nuvem de pontos de entrada referente aos dados da superfície sintética, o algoritmo

de detecção de planos foi aplicado. Os parâmetros: número de divisões e quantidade mínima de pontos da ferramenta não foram alterados durante os testes, e foram mantidos em $\eta = 20$ e $\delta = 100$. Os valores das variáveis da função f utilizadas para gerar os dados de entrada foram $\rho = 0,8$ e $\psi = 1,5$. O grau de ajuste m foi modificado aumentando a precisão dada ao algoritmo. A Figura 17 mostra os resultados obtidos para $m \in \{4, 5, 6, 7\}$, lembrando que $m \geq 4$ foi indicado como um bom valor de ajuste por Fernández (FERNÁNDEZ, 2005). O tamanho da nuvem de pontos gerada para estes testes é de 180.000 pontos.

Figura 17: Testes realizados alterando o parâmetro de ajuste do plano m .



Fonte: Imagem produzida pelo autor

Como explicado nas Seções 4.3 e 4.3.2 as cores dos planos são baseados nos valores de ângulo de mergulho e direção da camada do plano. Os valores de direção da camada não fazem sentido nesta análise, pois os dados não são georeferenciados. Entretanto as cores, neste caso, demonstram como a inclinação do plano segue a tendência dos dados. Os planos de cores verde claro são planos de grande inclinação (próximos a 90°), já planos com cor verde escuro são planos com pouca inclinação (próximos a 0°).

As linhas amarelas indicam o centro de cada plano, e também correspondem ao vetor normal do plano. Com estas informações podemos ver as diferenças gradativas que ocorrem em planos vizinhos. Estas diferenças gradativas são melhor visualizadas quando o parâmetro de ajuste do plano é maior, forçando a detecção de planos menores com pontos mais alinhados.

Com os resultados obtidos da análise dos dados sintéticos, percebe-se como o parâmetro m de ajuste do plano (coplanaridade) altera a detecção de planos na nuvem de pontos. Na Figura 17 quando utiliza-se $m \geq 4$ grandes planos são encontrados. Na visualização os pontos que foram utilizados para formar os planos tem a mesma corloração, e com isso podemos perceber que alguns pontos fazem parte de uma curva na superfície. Este resultado é esperado já que o parâmetro (m) permite estas variações nos dados. Aumentando o ajuste dos pontos ao plano com $m \geq 7$ podemos ver que os planos encontrados são menores e em maior quantidade. Nas grandes variações da superfície percebe-se que os pontos que estão posicionados nas inflexões da cova, não foram considerados para nenhum plano.

5.2 Modelo digital de afloramento

Os testes realizados utilizando dados reais de um mapeamento de afloramento foram sediados pelo laboratório de sensoriamento remoto e cartografia digital da Universidade do Vale do Rio dos Sinos (LASERCA - UNISINOS). O equipamento utilizado para obtenção dos dados é o *LASER Scanner ILIRS 3D (Intelligent LASER Ranging & Imaging System)* (OPTECH INCORPORATED INDUSTRIAL & 3D IMAGING DIVISION, 2006), ilustrado na Figura 18. Serão apresentados algumas imagens com resultados visuais na aplicação construída e a comparação dos dados obtidos através do método de detecção automática de planos com os dados coletados em campo.

Figura 18: Equipamento ILRIS3D da Optech



Fonte: Imagem retirada de (OPTECH INCORPORATED INDUSTRIAL & 3D IMAGING DIVISION, 2006)

O *scanner* acompanha uma câmera digital acoplada que é utilizada para obter informações de cor para o modelo 3D. A Figura 19 apresenta o afloramento mapeado e a nuvem de pontos renderizada.

Uma aplicação foi contruída para visualização dos dados da nuvem de pontos do afloramento estudado, conforme descrito na Seção 4.3. Em campo alguns planos foram medidos com bússola e clinômetro, obtendo os dados de direção da camada e ângulo de mergulho, respectivamente. A Figura 20 mostra um dos planos medidos e que será o foco de estudo.

A coleta das informações de orientação do plano foram realizadas por três profissionais realizando dez medições cada, assim totalizando trinta medições. A Tabela 1 apresenta as informações consolidadas desta coleta. Estes serão os valores de direção da camada (*DD*) e ângulo de mergulho (*DA*) utilizados para comparação com os valores calculados para os planos encontrados pelo algoritmo.

	<i>DD</i>	<i>DA</i>
<i>Média</i>	88,05°	86,23°
σ	1,24°	1,33°

Tabela 1: Dados medidos em campo do plano estudado

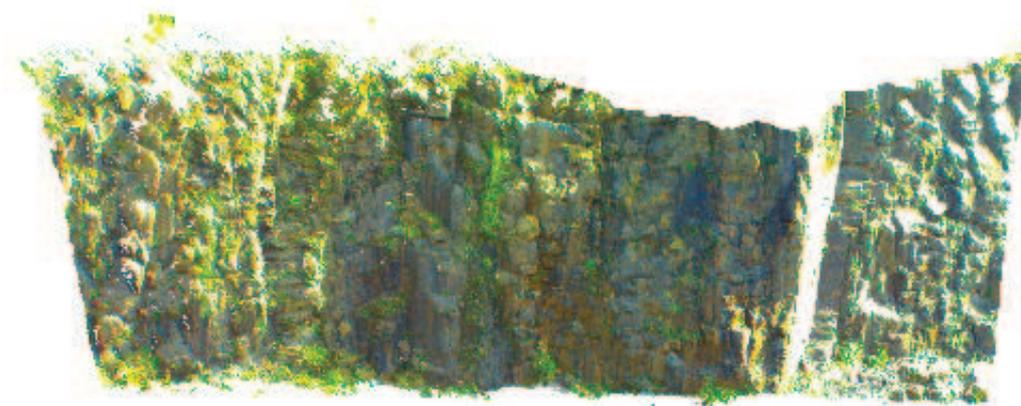
Foram realizados três testes para comparação dos dados de campo apresentados na Tabela 1 com os resultados obtidos pelo algoritmo. O primeiro teste visa validar o comportamento do

Figura 19: Afloramento localizado na pedreira Incopel (a), e modelo digital obtido do mapeamento do afloramento, visualizada na aplicação construída (b).



Fonte: Imagem cedida por LASERCA/UNISINOS

(a)



Fonte: Imagem produzida pelo autor

(b)

Figura 20: Plano de estudo



Fonte: Imagem cedida por LASERCA/UNISINOS

algoritmo com um conjunto pequeno de dados. Para isso foram extraídos, da nuvem de pontos, somente os pontos referentes ao plano estudado. O segundo teste faz um corte no modelo digital para avaliar o comportamento do algoritmo em um conjunto maior de informações. O terceiro teste avalia a execução do algoritmo quando aplicado ao modelo digital de afloramento como um todo.

5.2.1 Primeiro teste: plano isolado

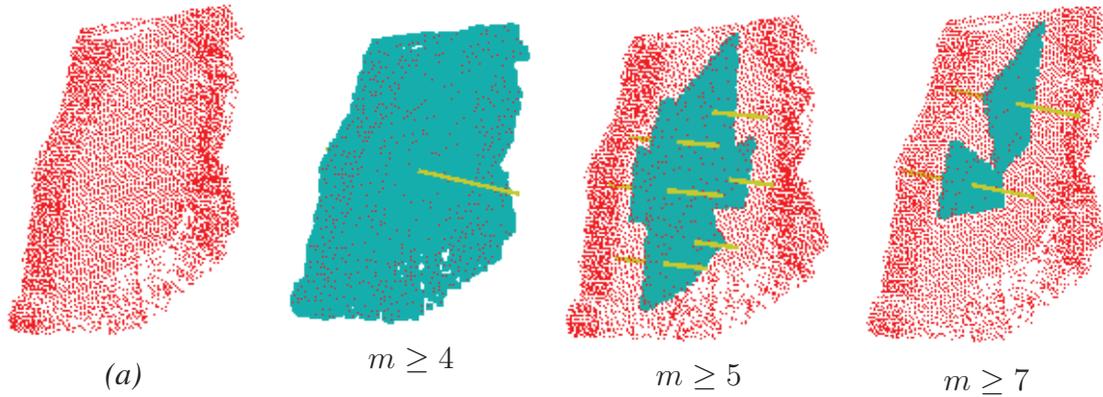
O primeiro teste realizado tem como objetivo validar os valores de ângulo de mergulho e direção da camada, calculados pelo algoritmo, isolando somente o plano estudado. Com esta validação controlada é esperado que o plano seja encontrado e que os valores calculados sejam próximos dos medidos em campo.

O algoritmo foi executado quatro vezes variando o valor do parâmetro m entre 4 e 5. A Figura 21 apresenta os resultados obtidos para cada execução, Figura 21 (a) são os pontos isolados da nuvem de pontos referentes ao plano estudado. O resultado para $m \geq 6$ não são apresentados, pois são os mesmo de $m \geq 5$.

Como salientado durante a descrição do algoritmo na Seção 4.1, diversos planos foram detectados na área estudada, como mostra a Figura 21. Nos casos onde mais de um plano foi calculado, os valores de ângulo de mergulho e direção da camada dos planos foi somado e uma média foi calculada, da mesma forma que foi realizado com os dados medidos em campo. A Tabela 2 apresenta os dados consolidados desta análise, e a Tabela 3 mostra as diferenças existentes entre os valores estimados e os medidos em campo.

Na Tabela 3 percebe-se que para $m \geq 5$ houve uma grande diferença no valor de direção da camada. Isso se deve pela natureza do plano, pois o mesmo não é perfeitamente plano. As pequenas variações que não são percebidas em medidas de campo podem ser avaliadas no

Figura 21: Análise dos pontos que formam o plano



Fonte: Imagem produzida pelo autor

	<i>DD</i>	<i>DA</i>	Número de Planos
$m \geq 4$	87,99°	87,84°	1
$m \geq 5$	83,83° $\sigma = 1,39$	85,49° $\sigma = 0,42$	6
$m \geq 7$	84,39° $\sigma = 1,36$	85,60° $\sigma = 0,62$	2

Tabela 2: Resultados dos testes executados sobre os dados do plano isolado.

	ΔDD	ΔDA
$m \geq 4$	-0,06°	1,61°
$m \geq 5$	-4,21°	-0,73°
$m \geq 7$	-3,65°	-0,63°

Tabela 3: Comparação entre os valores de orientação do plano calculados através da nuvem de pontos e os valores medidos em campo.

modelo digital.

5.2.2 Segundo teste: corte no modelo

Para realizar o segundo teste, a nuvem de pontos do modelo digital de afloramento foi corada de forma que contenha o plano estudado e outros possíveis planos a serem detectados. O objetivo desta seleção é avaliar como o método se comporta na identificação do plano estudado em uma nuvem de pontos com uma quantidade de pontos maior. A Figura 22 ilustra a comparação da imagem da pedra com o corte realizado na nuvem de pontos para os testes.

O mesmo processo realizado para os pontos do plano isolado foi aplicado ao corte na nuvem de pontos. Em todos os casos o plano estudado foi detectado e os valores de orientação do plano foram calculados. A Tabela 4 contém as informações obtidas durante os testes. Da mesma forma que no teste anterior, quando mais de um plano foi medido a média dos valores é exibida na tabela. A Tabela 5 apresenta as diferenças entre os valores calculados e os medidos em

Figura 22: Afloramento localizado na pedreira Incopel (a), e modelo digital obtido do mapeamento do afloramento (b).



Fonte: Imagem cedida por LASERCA/UNISINOS

(a)



Fonte: Imagem produzida pelo autor

(b)

campo.

	<i>DD</i>	<i>DA</i>	Número de Planos
$m \geq 4$	$84,45^\circ \sigma = 1,85$	$87,20^\circ \sigma = 1,39$	4
$m \geq 5$	$84,69^\circ \sigma = 1,03$	$85,71^\circ \sigma = 0,94$	6
$m \geq 7$	$84,73^\circ \sigma = 1,45$	$85,18^\circ \sigma = 0,31$	2

Tabela 4: Resultados dos testes executados sobre os dados do corte realizado no afloramento.

	ΔDD	ΔDA
$m \geq 4$	$-3,77^\circ$	$0,97^\circ$
$m \geq 5$	$-3,36^\circ$	$-0,51^\circ$
$m \geq 7$	$-3,32^\circ$	$-1,04^\circ$

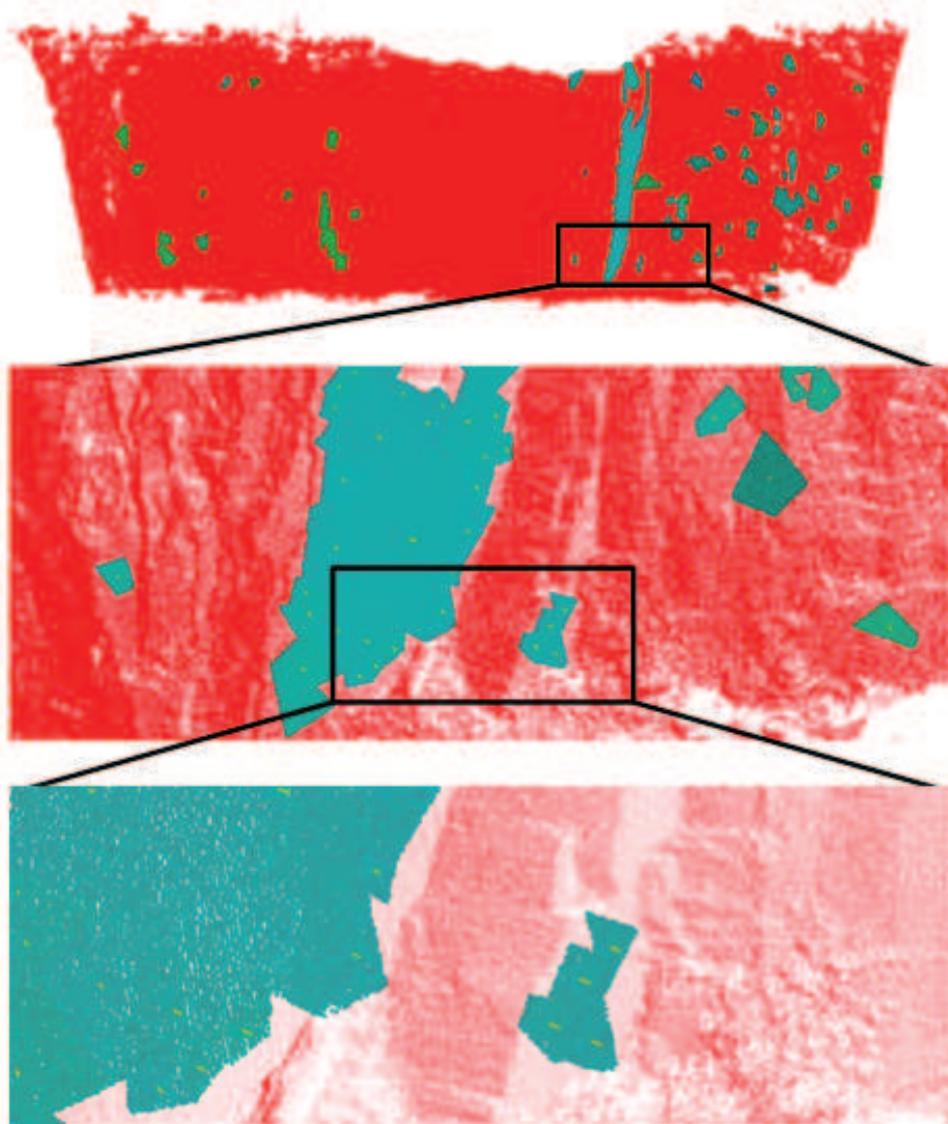
Tabela 5: Diferença entre os valores de orientação do plano calculados através da nuvem de pontos e os valores medidos em campo, para região estudada.

5.2.3 Terceiro teste: modelo completo

No terceiro testes, toda a nuvem de pontos obtida através do mapeamento do afloramento foi processada utilizando o algoritmo de detecção automática de planos. A nuvem de pontos contém 6.115.621 pontos, e durante os testes o parâmetro de ajuste do plano foi alterado.

A Figura 23 ilustra o resultado do processamento para $m \geq 6$.

Figura 23: Planos gerados através de detecção automática sobre o afloramento. Os parâmetros utilizados são $m \geq 6$, $k \leq 0.8$, $\eta = 20$ e $\delta = 100$.



Fonte: Imagem produzida pelo autor

A Tabela 8 apresenta os dados encontrados para o plano estudado. Da mesma forma, a Tabela 7 apresenta os dados de comparação dos valores de orientação do plano calculado e medido em campo.

	<i>DD</i>	<i>DA</i>	Número de Planos
$m \geq 5$	84, 22° $\sigma = 0, 49$	85, 44° $\sigma = 0, 63$	3
$m \geq 6$	84, 17° $\sigma = 0, 42$	85, 43° $\sigma = 0, 51$	4
$m \geq 7$	84, 39° $\sigma = 0, 53$	85, 43° $\sigma = 0, 51$	4

Tabela 6: Resultados dos testes executados sobre os dados do afloramento.

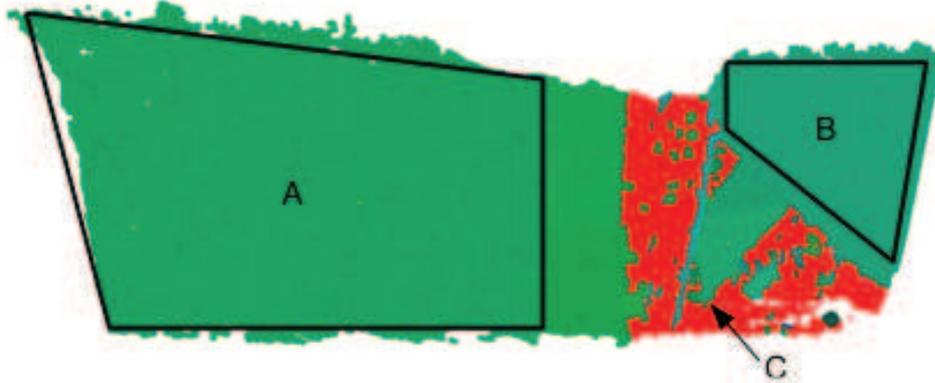
	ΔDD	ΔDA
$m \geq 5$	$-3,82^\circ$	$-0,79^\circ$
$m \geq 6$	$-3,87^\circ$	$-0,79^\circ$
$m \geq 7$	$-3,62^\circ$	$-0,79^\circ$

Tabela 7: Diferenças entre os valores de orientação do plano calculados através da nuvem de pontos e os valores medidos em campo.

Não são apresentados os dados obtidos para $m \geq 4$ pois, somente dois grandes planos que dividiam o afloramento foram encontrados. Esta imprecisão se dá ao fato da característica, neste caso, pouco restritiva imposta pelo valor do parâmetro m . Esta configuração permite que planos com muitas ondulações e diferenças nos pontos de origem sejam criados.

Este mesmo caso ocorre para $m \geq 5$, onde em partes da nuvem de pontos, grandes planos são encontrados, e sabe-se que pequenos planos estão contidos naquela área. A Figura 24 apresenta esta situação, onde os planos **A** e **B** foram calculados e satisfazem as propriedades estabelecidas para detecção de planos.

Figura 24: O processamento dos dados do afloramento para $m \geq 5$ encontrou grandes planos em áreas com deformidades, mas que satisfazem os parâmetros do algoritmo.



Fonte: Imagem produzida pelo autor

O plano **C** é o plano estudado, que foi encontrado durante o processamento com os dados descritos na Tabela 8. A área do afloramento onde se encontra este plano contém diversas fraturas, fazendo com que a variação dos pontos não satisfaça os parâmetros de detecção, e assim, mais divisões dos dados ocorram.

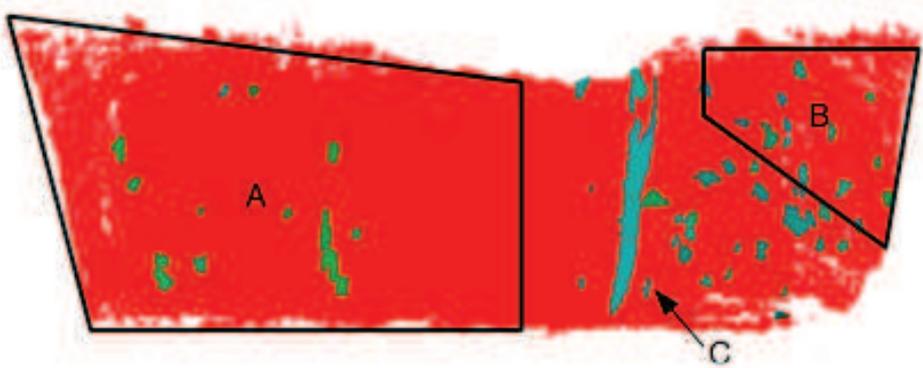
Para encontrar planos menores contidos nas áreas dos planos **A** e **B**, é necessário aumentar a precisão do algoritmo para detecção dos planos. Quando a nuvem de pontos do afloramento é processada com $m \geq 6$, planos menores são encontrados nas áreas onde antes existiam dois grandes planos Figura 25.

Com uma maior divisão dos dados foram detectados 29 planos na área do plano **A** e 17 planos na área do plano **B**.

	m	k	DD	DA
<i>Plano A</i>	5,26	0,19	26,96°	78,57°
<i>Plano B</i>	5,32	0,20	46,12°	79,01°

Tabela 8: Dados dos planos **A** e **B** encontrados durante a detecção dos planos. Os dois planos satisfazem as características estabelecidas para o algoritmo.

Figura 25: Planos encontrados na nuvem de pontos do afloramento quando o parâmetro de ajuste dos pontos ao plano foi aumentado.



Fonte: Imagem produzida pelo autor

Com os resultados apresentados é possível perceber que o algoritmo proposto atinge os objetivos deste trabalho. Com a possibilidade de variar os valores de ajuste do plano é possível obter maior detalhamento como mostra a Figura 25.

Os resultados comparativos dos valores de ângulo de mergulho e direção da camada (Tabelas 3, 5 e 7) apresentam variações médias de $-0,78^\circ$ e $-3,71^\circ$, respectivamente. Indicando valores muito próximos aos valores medidos em campo. Estas diferenças também são encontradas no trabalho realizado com instrumentos de medição pelos geólogos, onde mais de uma medida dos valores de ângulo de mergulho e direção da camada são realizados (SOUZA, 2012). Estas são diferenças que indicam um bom resultado visto que estão dentro do esperado em estudos de campo.

6 CONCLUSÃO

Para realizar pesquisas sobre a estrutura terrestre, geólogos se submetem a situações de risco. A medição de estruturas planares em afloramentos é uma delas. Devido a natureza desta fonte de estudo, com dimensões de centenas de metros, tais pesquisas não são exploradas em toda sua plenitude.

A utilização de técnicas de sensoriamento remoto como TLS, auxiliam na obtenção de informações sobre a estrutura de afloramentos. Minimizando os riscos e impedimentos impostos nas pesquisas de campo. Contudo um novo ramo de estudo é criado, sendo necessário processar e interpretar os dados obtidos com o TLS.

Existem diversos métodos conhecidos para o cálculo de estruturas planares em modelos digitais de afloramentos a partir dos dados do TLS (NAGALLI; FIORI; NAGALLI, 2011; FENG et al., 2001; ALAN K. L. KWONG; WONG, 2007; FERNÁNDEZ, 2005). Estes métodos calculam os valores de orientação dos planos a partir de uma seleção prévia dos dados do TLS.

O presente trabalho propôs um método de detecção automática de planos, utilizando a técnica de Análise de Componentes Principais. O algoritmo foi criado com o objetivo de ter uma nuvem de pontos de entrada de afloramentos. Contudo, qualquer nuvem de pontos pode ser aplicada ao algoritmo, já que todos os parâmetros criados para a detecção de planos são parametrizados (Seção 4.3). Mesmo para diferentes afloramentos estes valores deverão ser estudados para obter os melhores resultados possíveis.

Como apresentado na Seção 5.2, onde diversos testes foram realizados com diferentes tamanhos da entrada (nuvem de pontos), os valores de orientação do plano de estudo foram todos coerentes entre si. Houve de maneira geral uma diferença média de $\pm 3^\circ$ no valor de direção da camada do plano (Tabelas 3, 5 e 7). Isso pode ser decorrente de algum ruído durante o georreferenciamento dos pontos ou até mesmo de uma maior precisão para os valores calculados em relação aos medidos com bússola.

A utilização do TLS para medição de planos em afloramentos pode ser mais precisa e segura do que medições manuais de campo. Obtida a nuvem de pontos o algoritmo sempre calculará os mesmos planos, ao contrário de medições manuais em campo (SOUZA, 2012). Contudo, o processo de mapeamento do afloramento se torna mais crucial nestas análises, visto que qualquer erro neste passo irá afetar qualquer estudo posterior.

Algumas melhorias ainda podem ser realizadas no aplicativo desenvolvido com o objetivo de flexibilizar ainda mais as análises realizadas sobre os modelos digitais de afloramento. Pode-se citar: permitir uma seleção prévia mais restrita da nuvem de pontos do modelo pelo usuário. Isso permite que seções do mapeamento que foram pouco amostradas ou seções onde existe o conhecimento prévio de que não há planos interessantes, possam ser excluídas do processamento.

Existem também planos extremamente pequenos com centímetros de dimensão, que mesmo em campo são difíceis de serem medidos. Estes planos podem ser pequenos o suficiente que

ficam impraticáveis para as medições com bússola e clinômetro. Quando isto ocorre uma extensão é criada com auxílio de um material plano e então a medida é realizada. Casos similares ocorrem no modelo digital. O geólogo pode ter conhecimento prévio de um plano que não foi detectado automaticamente pelo algoritmo, ou a informação não foi amostrada durante o processo de mapeamento. É interessante que a aplicação permita o cálculo de atributos de orientação de planos baseando-se também em dados informados pelo usuário.

Mesmo que o método proposto neste trabalho tenha demonstrado robustez, existem diversos casos que o mesmo pode falhar. Para isso o importante é fornecer subsídios ao usuário para decidir a melhor forma de calcular todos os dados necessários.

REFERÊNCIAS

- ALAN K. L. KWONG, H. K.; WONG, A. Use of 3D Laser Scanner for Rock Fractures Mapping. **FIG Working Week**, [S.l.], p. 14, 2007.
- BALTSAVIAS, E. P.; FAVEY, E.; BAUDER, A.; BOSCH, H.; PATERAKI, M. Digital surface modelling by airborne laser scanning and digital photogrammetry for glacier monitoring. **The Photogrammetric Record**, [S.l.], v. 17, n. 98, p. 243–273, 2001.
- BONABEAU, E.; DORIGO, M.; THERAULAZ, G. **Swarm intelligence: from natural to artificial systems**. [S.l.]: Oxford university press New York, 1999. v. 4.
- BUCKLEY, S. J.; ENGE, H. D.; CARLSSON, C.; HOWELL, J. A. Terrestrial laser scanning for use in virtual outcrop geology. **The Photogrammetric Record**, [S.l.], v. 25, n. 131, p. 225–239, 2010.
- BUCKLEY, S. J.; HOWELL, J.; ENGE, H.; KURZ, T. Terrestrial laser scanning in geology: data acquisition, processing and accuracy considerations. **Journal of the Geological Society**, [S.l.], v. 165, n. 3, p. 625–638, 2008.
- CHEN, G.; QIAN, S.-e. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. **Geoscience and Remote Sensing, IEEE Transactions on**, [S.l.], v. 49, n. 3, p. 973–980, 2011.
- DANDPAT, S. K.; MEHER, S. Performance improvement for face recognition using PCA and two-dimensional PCA. In: INTERNATIONAL CONFERENCE ON COMPUTER COMMUNICATION AND INFORMATICS (ICCCI), 2013., 2013. **Anais...** [S.l.: s.n.], 2013.
- DAULTREY, S. **Principal components analysis**. [S.l.]: Geo Abstracts Limited Norwich, 1976.
- DAVIS, G. H.; REYNOLDS, S. J. **Structural Geology of Rocks and Regions**. 2. ed. [S.l.]: John Wiley and Sons, INC., 1996.
- EIGEN 3.2.0 Documentation. 2012.
- FABUEL-PEREZ, I.; HODGETTS, D.; REDFERN, J. A new approach for outcrop characterization and geostatistical analysis of a low-sinuosity fluvial-dominated succession using digital outcrop models: upper triassic oukaïmeden sandstone formation, central high atlas, morocco. **AAPG bulletin**, [S.l.], v. 93, n. 6, p. 795–827, 2009.
- FENG, Q.; SJÖGREN, P.; STEPHANSSON, O.; JING, L. Measuring fracture orientation at exposed rock faces by using a non-reflector total station. **Engineering Geology**, [S.l.], v. 59, n. 1, p. 133–146, 2001.
- FERNÁNDEZ, O. Obtaining a best fitting plane through 3D georeferenced data. **Journal of structural geology**, [S.l.], v. 27, n. 5, p. 855–858, 2005.
- FRÖHLICH, C.; METTENLEITER, M. Terrestrial laser scanning—new perspectives in 3D surveying. **International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences**, [S.l.], v. 36, n. Part 8, p. W2, 2004.

GROTZINGER, J.; JORDAN, T. H.; PRESS, F.; SIEVER, R. **Understanding Earth**. 5. ed. [S.l.]: W. H. Freeman and Company, 2007.

HODGETTS, D. Laser Scanning and Digital Outcrop Geology in the Petroleum Industry: a review. **Marine and Petroleum Geology**, [S.l.], 2013.

JACKSON, J. A.; MEHL, J. P.; NEUENDORF, K. K. **Glossary of geology**. [S.l.]: Springer, 2005.

JACKSON, J. E. **A user's guide to principal components**. [S.l.]: Wiley-Interscience, 2005. v. 244.

JOLLIFE, I. T. **Principal Component Analysis**. [S.l.]: Springer verlag, 2002.

KASS, M. A.; LI, Y. Quantitative analysis and interpretation of transient electromagnetic data via principal component analysis. **Geoscience and Remote Sensing, IEEE Transactions on**, [S.l.], v. 50, n. 5, p. 1910–1918, 2012.

MORRIS, J. T.; PORTER, D.; NEET, M.; NOBLE, P. A.; SCHMIDT, L.; LAPINE, L. A.; JENSEN, J. R. Integrating LIDAR elevation data, multi-spectral imagery and neural network modelling for marsh characterization. **International Journal of Remote Sensing**, [S.l.], v. 26, n. 23, p. 5221–5234, 2005.

NAGALLI, A.; FIORI, A. P.; NAGALLI, B. Método para aplicação de escâner a laser terrestre ao estudo da estabilidade de taludes em rocha. **Revista Brasileira de Geociências**, [S.l.], v. 41, n. 1, p. 56–67, 2011.

NOKIA. **Qt 4.6 Whitepaper**. [S.l.]: Nokia Corporation, 2009. disponível em <http://qt.nokia.com/files/pdf/qt-4.4-whitepaper>.

OPTECH. **ILRIS-3D Operation Manual**. 0040170/Rev A. ed. 300 Interchange Way - Vaughan, Ontario - Canada: [s.n.], 2006. www.optech.ca.

PEDERSEN, S. I.; SKOV, T.; RANDEN, T.; SØNNELAND, L. Automatic fault extraction using artificial ants. In: **Mathematical Methods and Modelling in Hydrocarbon Exploration and Production**. [S.l.]: Springer, 2005. p. 107–116.

QIHONG, Z.; XING, X.; YOUYAN, Z.; YONG, Y.; YAN, H.; SONG, L. Digital outcrop modeling and geology information extraction based on ground-based Lidar. In: **AUDIO, LANGUAGE AND IMAGE PROCESSING (ICALIP), 2012 INTERNATIONAL CONFERENCE ON**, 2012. **Anais...** [S.l.: s.n.], 2012. p. 580–583.

RUSSELL, S. J.; NORVIG, P.; CANNY, J. F.; MALIK, J. M.; EDWARDS, D. D. **Artificial intelligence: a modern approach**. [S.l.]: Prentice hall Englewood Cliffs, 1995. v. 74.

SOUZA, M. K. de. **Laser scanner terrestre: uma ferramenta eficaz para medidas de estruturas geológicas em afloramentos**. 2012. Dissertação (Mestrado em Ciência da Computação) — Universidade do Vale do Rio dos Sinos, 2012.

STURZENEGGER, M.; STEAD, D. Close-range terrestrial digital photogrammetry and terrestrial laser scanning for discontinuity characterization on rock cuts. **Engineering Geology**, [S.l.], v. 106, n. 3, p. 163–182, 2009.

VAN GENECHTEN, B. **Theory and practice on Terrestrial Laser Scanning**: training material based on practical applications. [S.l.]: Universidad Politecnica de Valencia Editorial, 2008.

VOSSelman, G.; GORTE, B. G.; SITHOLE, G.; RABBANI, T. Recognising structure in laser scanner point clouds. **International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences**, [S.l.], v. 46, n. 8, p. 33–38, 2004.

WANG, R.; QIAN, X. **Openscenegraph 3.0**: beginner's guide. [S.l.]: Packt Publishing, Limited, 2010.

WOODCOCK, N. Specification of fabric shapes using an eigenvalue method. **Geological Society of America Bulletin**, [S.l.], v. 88, n. 9, p. 1231–1236, 1977.