UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E INFORMÁTICA INDUSTRIAL

JOSÉ ROSA KUIASKI

SEGMENTAÇÃO DE MOVIMENTO POR FLUXO ÓTICO

DISSERTAÇÃO

CURITIBA

2012

JOSÉ ROSA KUIASKI

SEGMENTAÇÃO DE MOVIMENTO POR FLUXO ÓTICO

Dissertação apresentada para o Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito parcial para a obtenção do grau de "Mestre em Ciências" – Área: Engenharia de Computação.

Orientador: Prof. Dr. Hugo Vieira Neto

CURITIBA

2012

Dados Internacionais de Catalogação na Publicação

K96 Kuiaski, José Rosa
Segmentação de movimento por fluxo ótico / José Rosa Kuiaski. – 2012.
90 f. : il. ; 30 cm

Orientador: Hugo Vieira Neto.
Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. Curitiba, 2012.
Bibliografia: f. 88-90.

1. Percepção visual do movimento. 2. Processos percepto-motores –Testes. 3. Movimento. 4. Algoritmos. 5. Simulação (Computadores). 6. Engenharia elétrica – Dissertações. I. Vieira Neto, Hugo, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial. III. Título.

CDD (22. ed.) 621.3

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
Câmpus Curitiba

Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial

Título da Dissertação Nº 604

"Segmentação de Movimento por Fluxo Ótico"

por

José Rosa Kuiaski

Esta dissertação foi apresentada como requisito parcial à obtenção do grau de MESTRE EM CIÊNCIAS – Área de Concentração: Engenharia da Computação, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial – CPGEI – da Universidade Tecnológica Federal do Paraná – UTFPR – Câmpus Curitiba, às 14h do dia 24 de agosto de 2012. O trabalho foi aprovado pela Banca Examinadora, composta pelos doutores:

Prof. Hugo Vieira Neto, Dr.
(Presidente – UTFPR)

Prof. Eduardo Todt, Dr.
(UFPR)

Prof. Marcelo Victor Wüst Zibetti, Dr.
(UTFPR)

Prof. Gustavo Benvenutti Borba, Dr.
(UTFPR)

Visto da coordenação:

Prof. Ricardo Lüders, Dr.
(Coordenador do CPGEI)

AGRADECIMENTOS

Deo, pro univérsis benefíciis tuis.

À minha esposa, Michelle Kuiaski, pelo apoio moral, companheirismo e paciência.

Aos meus pais, meus primeiros e eternos professores!

Ao meu irmão, Diogo Rosa Kuiaski, por me mostrar o caminho das pedras!

Ao professor Hugo Vieira Neto, por acreditar no meu trabalho, pela paciência, sabedoria e amizade!

Ao professor Marcelo Zibetti, pelas palavras certas na hora certa, que definiram o rumo desta pesquisa.

Aos professores Eduardo Todt e Gustavo Borba, pela atenção e pelo direcionamento.

Aos meus amigos do CPGEI: Thiago Bassani, Charles Fung, Liane Lubrigati e Alisson, pela disposição, pelo incentivo e por tornarem o trabalho mais leve e agradável.

Aos meus amigos da Denke pela tolerância e apoio.

“If we knew what it was we were doing, it would not be called research, would it?”

Albert Einstein

RESUMO

KUIASKI, José Rosa. Segmentação de Movimento por Fluxo Ótico. 2012. 90 f. Dissertação – Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2012.

A percepção de movimento é uma característica essencial à sobrevivência de diversas espécies. Na natureza, é através do movimento que uma presa percebe a chegada de um predador e decide em que direção deve fugir, bem como o predador detecta a presença de uma presa e decide para onde atacar. O Sistema Visual Humano é mais sensível a movimento do que a imagens estáticas, sendo capaz de separar as informações de movimento originadas pela movimentação própria das informações de movimento de objetos animados no ambiente. A Teoria Ecológica de Gibson (1979) provê uma base para o entendimento de como esse processo de percepção ocorre e estende-se com o conceito do que chamamos de campo vetorial de Fluxo Ótico, através do qual se representa computacionalmente o movimento. O objetivo principal deste trabalho é procurar reproduzir computacionalmente esse comportamento, para possíveis aplicações em navegação autônoma e processamento de vídeo com movimentação desconhecida. Para isso, vale-se das técnicas de estimação de Fluxo Ótico presentes na literatura, tais como as propostas por Lucas e Kanade (1981) e Farneback (1994). Em um primeiro momento, avalia-se a possibilidade de utilização de uma técnica estatística de separação cega de fontes, a chamada Análise de Componentes Independentes, tomando como base o trabalho de Bell e Sejnowski (1997), na qual se mostra que tal análise aplicada em imagens fornece filtros de bordas. Depois, avalia-se a utilização do Foco de Expansão para movimentos translacionais. Resultados experimentais demonstram uma maior viabilidade da abordagem por Foco de Expansão.

Palavras-chave: Análise de Componentes Independentes, Fluxo Ótico, Movimento, *Egomotion*

ABSTRACT

KUIASKI, José Rosa. Motion Segmentation through Optical Flow. 2012. 90 f. Dissertação – Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2012.

Motion Perception is an essential feature for the survival of several species. In nature, it is through motion that a prey perceives the predator and is able to decide which direction to escape, and the predator detects the presence of a prey and decides where to attack. The Human Visual System is more sensitive to motion than to static imagery, and it is able to separate motion information due to egomotion from that due to an animated object in the environment. The Ecological Theory of Gibson (1979) provides a basis for understanding how this processes of perception occurs, and leads to the concept of what we call the vector field of Optical Flow, through which computational motion is represented. The main objective of this work is to try to reproduce computationally this behaviour, for possible applications in autonomous navigation and video processing with unknown self-motion. For this, we use some Optical Flow estimation techniques, as those proposed by Lucas and Kanade (1981) and Farneback (1994). At first, we assess the possibility of using a statistical technique of blind source separation, the so-called Independent Component Analysis, based on the work of Bell and Sejnowski (1997), which demonstrates that this technique, when applied to imagery, provides edge filters. Then, we assess the use of the Focus of Expansion to translational motion. Experimental results show the second approach, using the Focus of Expansion, is more viable than through Independent Component Analysis.

Keywords: Independent Component Analysis, Optical Flow, Motion, Self-Motion

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

SVH	Sistema Visual Humano
ACI	Análise de Componentes Independentes
ICA	Independent Component Analysis
CI	Componente Independente
ACP	Análise de Componentes Principais
AOA	Arranjo Ótico de Ambiente
PI	Ponto de Interesse
FOE	Foco de Expansão
ZCA	Zero-phase Component Analysis

SUMÁRIO

1 INTRODUÇÃO

Movimento é a variação da posição espacial de um corpo no tempo. A percepção do movimento no meio ambiente é uma das características essenciais à sobrevivência de todas as espécies que se guiam pela visão, como é o caso do ser humano. Por exemplo, uma presa precisa identificar informação a respeito do movimento do predador e deve tomar o rumo contrário para sobreviver e, embora audição, olfato e tato tenham grande importância para identificar e alertar sobre perigos, é a visão e a capacidade de percepção de movimento que provê informações de longo alcance e a noção de localização.

Palmer (1999) diz que essa necessidade tornou o Sistema Visual Humano (SVH) mais sensível a movimento do que a imagens puras. Entretanto, ainda não há um consenso sobre a forma como o SVH processa as informações de movimento que capta do ambiente. Alguns estudos, como Longuet-Higgins e Prazdny (1980) consideram que as cenas tridimensionais do cotidiano são projetadas no plano da retina como uma imagem bidimensional através de transformadas projetivas, em uma taxa de aproximadamente 60 imagens por segundo.

A figura 1 mostra uma representação do SVH. Entender o SVH é importante pois existe uma analogia entre o mesmo, os sistemas de captação de imagens e vídeos e os sistemas de processamento. Como será visto no capítulo 2, a dinâmica entre o sistema de captação e processamento representado pelo SVH é a base de várias linhas da psicologia de percepção, inclusive a Ecológica de Gibson (1966) na qual este trabalho se baseia.

A primeira estrutura do SVH é o olho, conforme a figura 2. Sua função é receber os raios de luz refletidos no ambiente ao redor do observador e convertê-los em sinais elétricos que podem ser transmitidos pelo nervo ótico. A luz que chega aos olhos passa primeiramente por uma cavidade transparente, a córnea, que é preenchida por um líquido igualmente transparente chamado *humor aquoso*.

Após a córnea, existe uma membrana opaca chamada *íris*, que contém uma pequena abertura, a *pupila*, que regula a quantidade de luz que alcança o interior do olho. Basicamente, a pupila dilata quando a luminosidade é baixa, permitindo que mais luz seja capturada, e se

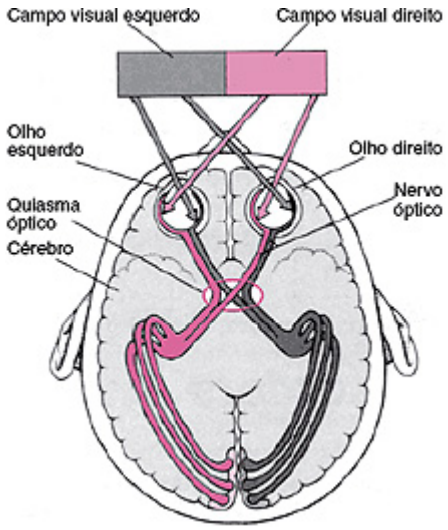

Figura 1: Representação do Sistema Visual Humano.

Fonte: Manual Merck de Informação Médica (MERCK, 2010)

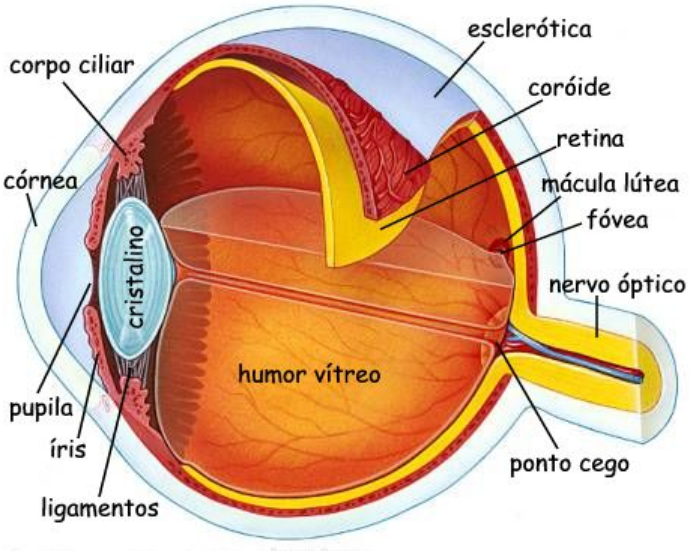

Figura 2: Olho humano.

Fonte: Retina (2009)

contrai quando a iluminação é alta. Palmer (1999) argumenta que essas dilatações e contrações respondem a fatores emocionais e psicológicos. Isso fornece indícios de que a percepção visual - e consequentemente de movimento - não é puramente fisiológico.

Uma vez que a quantidade correta de luz entra no olho, ela passa através do cristalino, uma lente com forma controlada que direciona os feixes de luz para a retina. A retina é uma membrana onde se encontram os *fotorreceptores*. Estes são estruturas que traduzem a energia luminosa em sinais neuronais e podem ser classificados segundo a sua forma: cones e bastonetes. A função de ambos é a mesma, mas os bastonetes são mais sensíveis à luz monocromática e encontram-se ao longo de toda a retina, menos em seu centro. Os cones são em menor número, sendo responsáveis pela visão em cores em condições normais de iluminação e concentram-se em uma região pequena da retina, chamada *fóvea*. A fóvea abriga a maior densidade de fotorreceptores e, por isso, apresenta visão espacial e em cores mais aguçada.

A partir da retina, não há consenso de como a informação luminosa se transforma em significado. Sob a ótica da corrente psicológica da *Gestält* (WEITHEIMER, 1923), pode-se inferir que a retina recebe imagens e que existe um processo cognitivo de ordem mais elevada, prossivelmente no córtex, que processa tais imagens para perceber movimento. Este tópico será discutido na seção 1.1, como uma peça importante na justificativa deste trabalho.

A questão de percepção de movimento leva a uma justificativa maior: a simples movimentação de um indivíduo leva à variação do padrão luminoso que alcança o plano da retina e esse processo cognitivo perceberia essa variação igualmente como movimento, definido como movimento próprio ou *egomotion*. A presença de um agente animado externo somaria as variações dos padrões de luz que chegam na retina, gerando um padrão composto. Entretanto, o cérebro humano é capaz de separar tais padrões e identificar o agente causador de cada um. A essa habilidade, será dado o nome de *segmentação de movimento*, o qual deseja-se reproduzir computacionalmente para aplicações de navegação autônoma e processamento de video.

Não é o objetivo deste trabalho responder como essa operação é realizada no cérebro humano, mas testar a hipótese de que ela pode ser realizada através de uma técnica estatística chamada de Análise de Componentes Indepedentes (HYVÄRINEN et al., 2001), ACI. Por questões de tradição, ao decorrer deste trabalho, será usada a sigla em ingles ICA.

As técnicas de ICA são apresentadas como formas para a realização de separação cega de fontes (COMON, 1994), nas quais parte-se do princípio de que as fontes, ou *componentes indepententes*, CI, são variáveis aleatórias estatisticamente independentes, conforme será abordado no capítulo 4.

1.1 MOTIVAÇÃO

Movimento, em nível de visão, tem sido estudado desde o final do século XIX. A primeira abordagem era psico-física e iniciou-se com o movimento da *Gestält*, que veio com a ideia de que a percepcão é mais do que somente estímulos em elementos sensoriais, mas existe um processo cognitivo de alta ordem sobre esses estímulos. Esse processo cognitivo pode ser encarado como um elemento a mais no espaço perceptorial. Então, a percepção de um único elemento é condicionada à percepção do todo. Por exemplo, só seria possível perceber uma nota musical desafinada dentro de uma melodia quando já se conhece a melodia inteira. Weitheimer pode ser apontado como o primeiro psicólogo a estudar visão. Em seu trabalho (WEITHEIMER, 1923), ele demonstra que existe um processo cognitivo que junta estímulos semelhantes através de uma relação que ele chamou de *Fator de Similaridade*.

De acordo com Marr (1982), a questão de Fluxo Ótico foi considerada primeiramente por James J. Gibson (GIBSON, 1966). Ele considerou que a percepção utiliza dados sensoriais passados conjuntamente com informação atual. Para o SVH, isso representa a capacidade de correlacionar localizações especiais através do tempo e classificar as características do ambiente como variáveis ou constantes.

A abordagem sugerida por Gibson (1966, 1977, 1979) é interessante quando se lida com a estimação computacional de Fluxo Ótico, conforme será abordado no capítulo 3, pois a mesma considera que o campo vetorial de Fluxo Ótico possui em si todas as informações necessárias para perceber movimento. Isso implica que não é necessária informação estereoscópica para inferir movimento relativo e que o problema de correspodência de Ullman (1979) poderia ser resolvido no tempo, em vez de no espaço.

Os fundamentos matemáticos da abordagem de Gibson vieram em 1980 com Longuet-Higgins e Prazdny (1980). Eles representaram matematicamente o movimento e o campo de Fluxo Ótico sobre o plano da retina uma vez que as velocidades de translação e rotação são conhecidas. De acordo com os autores, o campo de Fluxo Ótico é a soma vetorial das componentes de translação e rotação. Essa estrutura tridimensional é projetada em um espaço bidimensional que pode ser entendido como o plano de retina.

No escopo das aplicações de Visão Computacional, o plano da retina é visto como o plano de imagem, as coordenadas espaciais em metros como coordenadas em *pixels* e o campo de Fluxo Ótico como a variação da posição de pixels entre dois quadros consecutivos de um vídeo. Com a utilização de microcomputadores, as pesquisas na área de processamento de movimento ganharam visibilidade e impulsão. Surgiram os algoritmos que compõem a base do

processamento de movimento, como Moravec (1980), Horn e Schunck (1980), Lucas e Kanade (1981), Farnebäck (2001).

Considerando um campo de Fluxo Ótico como um conjunto de vetores, com direções e amplitudes conhecidas, esses vetores devem descrever tanto o movimento próprio da câmera quando o de um objeto animado independente e ambos podem ser parametrizados independentemente.

Em 1983, Herault et al. (1985) propuseram uma solução para a separação cega de fontes baseada em independência estatística, ICA, como uma extensão dos princípios de descorrelação estatística da Análise de Componentes Principais, ACP. A partir daí, muitos autores abordaram a separação cega de fontes com base em independência. Comon (1994) definiu formalmente o problema de ICA e o seu modelo atual. Bell e Sejnowski (1997) sugeriram que o processo de ICA agiria como um redutor de redundâncias e que a sua aplicação em imagens naturais resultaria em um conjunto de filtros visuais de bordas. Essa sugestão é um indicativo de que as bordas podem ser identificadas no córtex do SVH como um processo parecido ao de ICA.

Partindo-se das ideias de que o campo de Fluxo Ótico representa a variação do padrão luminoso na retina e que por si só contém as informações necessárias para a percepção de movimento e que o SVH realiza um processo de retirada de redundâncias que resulta em filtros de bordas, espera-se estender esse conceito no tempo e que as técnicas de ICA possam, de alguma forma, ser utilizadas para separar as informações de movimento próprio da câmera e de objetos animados independentes no meio.

Para isso, no capítulo 2 será apresentada a teoria Ecológica de Gibson, que trata a questão o Fluxo Ótico como uma peça chave e suficiente para a percepção de movimento. No capítulo 3, serão apresentadas as técnicas computacionais de estimação dos campos de Fluxo Ótico, de obtenção de pontos de interesse e de classificação dos mesmo por descritores SURF (BAY et al., 2008). O capítulo 4 apresenta a teoria que embasa as técnicas de ICA e mostra um exemplo simples da sua utilização. O capítulo 5 trata da questão da utilização das técnicas de ICA para o caso do Fluxo Ótico estimado. Alguns exemplos são mostrados e uma discussão das limitações das técnicas são feitas. Uma técnica alternativa de segmentação de movimento é apresentada no capítulo 6, bem como o processo de criação de uma base de dados sintética e implicações qualitativas inerentes ao processamento de vídeo baseado em *pixels*. Por fim, no capítulo 7, são salientadas as conclusões obtidas no decorrer deste trabalho e são propostos trabalhos futuros.

1.2 OBJETIVOS

1.2.1 OBJETIVO GERAL

O objetivo principal deste trabalho é segmentar o campo vetorial de Fluxo Ótico em duas componentes: uma relativa ao movimento próprio do observador ou da câmera, o conhecido *egomotion*; outra relativa à movimentação de um ou mais agentes independentes no meio.

1.2.2 OBJETIVOS ESPECÍFICOS

Para que o objetivo geral seja alcançado, os seguintes objetivos específicos foram traçados.

- Avaliar a possibilidade de segmentação dos vetores de Fluxo Ótico por técnicas de Análise de Componentes Independentes.
- Avaliar a possibilidade de segmentação dos vetores de Fluxo Ótico pela estimação do Foco de Expansão.

2 TEORIA ECOLÓGICA DE GIBSON

A importância da Teoria Ecológica de Gibson para este trabalho é que a mesma dá as bases cognitivas do processo de visão nos seres humanos. Parte-se da premissa que a partir dessas bases, o mesmo comportamento tende a ser replicado em um ambiente computacional. A Teoria Ecológica, em si, opõe-se às linhas tradicionais da psicologia. Por isso, primeiramente será apresentada uma linha histórica de como e porquê Gibson desenvolveu esse novo paradigma. Uma possível abordagem para esta pesquisa é a utilização das teorias psicológicas que seguem como forma de procurar uma nova forma de processamento de movimento. No contexto atual, tais teorias são fatores motivacionais, mais do que teóricos.

O início do século XX foi marcado pela crescente onda de experimentalismo com psicologia animal. Gatos e ratos brancos eram os objetos de estudo de comportamento e alguns experimentos, como o do enigma do gato em caixas de Edward Lee Thorndike (1898) e a navegação de ratos brancos em labirintos de Williard Small (1900), ficaram famosos e serviram de base para uma geração de psicólogos experimentalistas. Nesse contexto, encontramos Ivan Pavlov, psicólogo premiado com o nobel de fisiologia de 1904. Os estudos da psicologia baseados em comportamento animal, principalmente os trabalho de Pavlov deram material a um novo paradigma da psicologia, o "behaviorismo", ou psicologia comportamental, que pregava que o comportamento humano pode ser explicado por efeitos de estímulo e resposta a estes.

Ao mesmo tempo, desenvolvia-se uma outra linha de psicologia, na Alemanha, chamada de *Gestält*. A "Psicologia da Forma", cujo paradigma baseia-se na presença de processos cognitivos que dão sentido à realidade na qual o indivíduo está inserido, contrapondo-se ao comportamentalismo clássico.

Gibson iniciou seus trabalhos na psicologia experimental na época do pleno florescimento dessas duas correntes psicológicas. Diferentemente de seus colegas behavioristas e gestaltistas, Gibson trabalhava com comportamento e percepção humana, o que abriu caminho para um novo paradigma da psicologia comportamental. Segundo Gibson, a psicologia até então carecia de maior caráter científico e de uma influência menor do associativismo do século

XIX e do racionalismo do século XVII. Era necessário um novo entendimento sobre o que é realidade e a sua crítica era que a psicologia científica nunca avançaria se as suas próprias bases não fossem revisadas e eventualmente refeitas. Para Gibson, o foco central do estudo da psicologia é o entendimento de "como"o indivíduo situa a si mesmo no ambiente; uma abordagem "ecológica" da psicologia (GIBSON, 1950).

Antes de nos aprofundarmos na Teoria Ecológica de Gibson, porém, é interessante entendermos mais a fundo o conceito da psicologia gestaltista, pois Gibson era essencialmente um psicólogo da *Gestält*. A sua crítica às linhas da psicologia da época (inclusive de algumas facetas da *Gestält*) partiram desse ponto e dessa formação.

2.1 *GESTÄLT* E A PERCEPÇÃO INDIRETA

A ideia que permeia a *Gestält* é a consideração de que existe um processo de organização dentro do cérebro humano. Esse processo organizacional é holístico, no qual primeiro se tem consciência do todo e somente após, se tem a consciência das partes que individualmente compõem o todo. Isso implica dizer que não só os órgãos sensoriais captam estímulos do meio, mas que existe um processo cognitivo que processa essa informação e dá sentido a elas.

Logo, a noção de realidade está no processo de identificar o todo e na posterior quebra do todo em partes que fazem sentido. Essa definição surgiu a partir das observações de Wertheimer (1913) sobre o efeito estroboscópico em um brinquedo representando a movimentação de um trem. Quando as fotos eram colocadas sobre uma sucessão rápida de flashes, dava-se o efeito de movimento contínuo, efeito que posteriormente viria a ser chamado de *movimento aparente*. Essa primeira hipótese conta que o que se vê não é puramente a soma das imagens, mas uma ilusão de ótica sobre o todo.

A partir dessa definição, Wertheimer propôs uma série de leis, ou princípios organizacionais, que compõem as *Leis da Gestält*.

- Lei de Pragnanz ou Lei da Gravidez: ideia de que o cérebro humano tende a completar imagens segmentadas, dando a impressão de continuidade. Aqui, o termo "gravidez" é uma analogia semântica à gravidez humana, no qual um significado (o de continuidade) é gerado a partir de outro (a soma das partes). Observe a figura 3. Ela é constituida de uma série de linhas. Existe algum processo cognitivo humano que completa as linhas e dá sentido à imagem.

- Lei da Similaridade: Existe um processo organizacional que tende a agrupar objetos

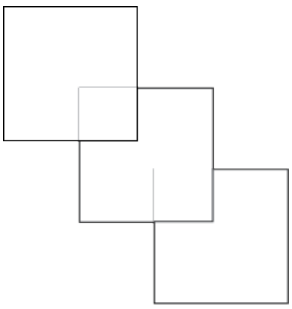

Figura 3: Figura demonstrativa da Lei de Pragnanz
Comumente chamada Lei da Gravidez.
Fonte: Autoria própria.

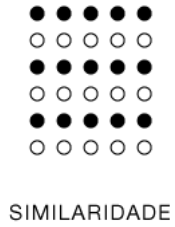

Figura 4: Figura demonstrativa da Lei de Similaridade
Objetos similares são agrupados.
Fonte: Autoria própria.

similares e esse agrupamento tende a fazer parte de um objeto maior (um "todo"ou um "gestalt"). Observe a figura 4. O cérebro tende a agrupar os objetos parecidos, no caso, círculos preenchidos fazem parte do mesmo grupo.

- Lei da Proximidade: Figuras próximas se agrupam. Observe a figura 5. A proximidade entre cada círculo define os agrupamentos, nesse caso, três.

- Lei da Simetria: Objetos simétricos têm sentido. Mesmo com a Lei da Proximidade, o fato de existir uma simetria se sobrepõe. A figura 6 demonstra o efeito.

Figura 5: Figura demonstrativa da Lei da Proximidade
Objetos próximos são agrupados.
Fonte: Autoria própria.

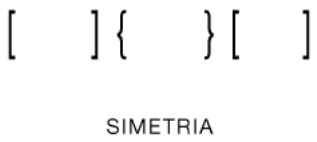

Figura 6: Figura demonstrativa da Lei da Simetria
Fonte: Autoria própria.

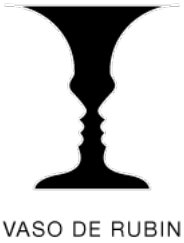

Figura 7: Vaso de Rubin
Fonte: Adaptação do site Illusioni Ottiche. http://www.illusioniottiche.net/

- Lei da Continuidade: A mente tende a dar continuidade a padrões, em casos de oclusão parcial, e dá sentido a uma imagem única e não a uma composição de imagens segmentadas.

Por fim, talvez o exemplo mais comum e mais significativo quando se fala das leis da *Gestält* é a noção de "Figura de fundo", de Edgard Rubin, conhecido como efeito "vaso de Rubin". Observe a figura 7. O significado da figura depende de qual referencial se assume para fundo e qual referencial se assume para a figura. Caso se considere o plano preto como fundo, vê se duas faces. Caso contrário, vê-se um vaso. De acordo com Rubin, é impossível ver ambas as imagens ao mesmo tempo.

Assumir um processo cognitivo ou organizacional leva a uma representação indireta de realidade. Até então, a ideia era a de que percepção é a construção de uma representação mental das informações que chegam nos órgãos receptivos; um processamento de informações sensoriais. Essa visão levou à conclusão de que a percepção humana é um processo de lógica e criatividade: ou seja, sem uma relação direta com o ambiente. Dessa forma, existiriam dois processos distintos até a percepção, segundo Helmholtz (1897):

- O estímulo: um processo puramente mecânico que envolve a recepção de informações nos órgãos sensoriais e o caminho pelos nervos;

- O mentalismo: o processo de tradução dessas informações em uma representação de realidade; em significados.

2.2 ABORDAGEM ECOLÓGICA E PERCEPÇÃO DIRETA DE REALIDADE

A base da abordagem Ecológica é a unificação funcional dos processos de estímulo e mentalismo, portanto a divisão entre corpo, órgãos sensoriais e nervosos, e mente é considerada errada (GIBSON, 1979). Gibson rejeitou a ideia de que a percepção seguisse um modelo de estímulo-resposta. Ao invés disso, já existiria no ambiente ao redor do observador um conjunto de informações prontas para serem percebidas por qualquer observador. Portanto, seriam informações "ecológicas" (GIBSON, 1979), em constraste à ideia de percepção baseada em imaginação e cognição simbólica.

Todos os seres estão sujeitos a informações ecológicas, todo o tempo. Entretanto, é necessário que o indivíduo observador esteja em busca dessas informações para que ocorra a percepção. Caso contrário, elas geram apenas sensações, não percepção (REED, 1989; GIBSON, 1950). Essa definição muda o objeto de estudo dos sentidos propriamente ditos para a habilidade humana de coleta de informações ecológicas através de sistemas perceptuais: uma conjunção de processos neurais e psicológicos com alto grau de capacidade adaptativa e com a finalidade de coletar informações do ambiente. A própria finalidade da percepção humana pode ser reinterpretada. O que antes era um processo de dar significado às informações coletadas, passa a ser uma forma de manter o indivíduo diretamente em contato com o ambiente.

Quando falamos de psicologia da visão, precisamos separar os conceitos físicos de tempo e espaço do conceito de ambiente. O que Gibson propôs é substituir, em nível psicológico, a ontologia tempo-espaço por uma ontologia persistência-mudança. Isso diz, sutilmente, que o os conceitos de espaço e tempo são abstratos e não são relevantes à percepção humana do meio. O Sistema Perceptual Humano não seria capaz de perceber o tempo ou o espaço diretamente, mas somente após ter contato com os objetos e com a suas mudanças ao longo do tempo. Entretanto, os conceitos de persistência e mudança estão sempre correlacionados. A persistência, principalmente a de longo prazo, é a que dá a noção de mundo e o suporte para a percepção. As mudanças carregam informações de percepção e são, efetivamente, os eventos ecológicos.

Como psicólogo da visão, as teorias de Gibson consideravam primordialmente a percepção visual. Uma das grandes contribuições de Gibson foi aplicar o conceito de persistência e mudança em nível da visão. Cada observador não apenas veria o ambiente de um ponto de

visão, mas sim veria um caminho de visão do meio. Um observador em movimento teria o seu caminho de visão constantemente em mudança. Entretanto, o conjunto de todos os caminhos de visão possíveis persistiria. O ambiente, como o conhecemos, seria composto pelo conjunto integral de todos os caminhos de visão possíveis. Assim, Gibson propôs o seu conceito de persistência-mudança como forma de resolver o dualismo corpo-mente que a doutrina psicológica até então pregava, quando se discriminava o ambiente individual (que está em constante mudança e que se interpola com o caminho de visão de outros observadores), do ambiente coletivo (que persiste).

Resta ainda uma questão a ser resolvida no que tange o ambiente coletivo. Como é possível que o ambiente seja o mesmo para todos os observadores, se é fisicamente impossível que dois deles estejam no mesmo lugar ao mesmo tempo? Logo, não seria possível que um ambiente fosse o mesmo para ambos. Nesse ponto, Gibson diferencia o que é visível em um determinado instante do que é puramente visível. Embora haja essa restrição física, dois observadores podem estar no mesmo lugar em tempos diferentes e, enquanto houver persistência do meio, cada observador pode explorá-lo. Essa consideração é a grosso modo a de que o ambiente engloba a todos os indivíduos observadores da mesma maneira que um observador individual.

Uma consideração interessante que Gibson tinha do ambiente individual é que ele provê sensações. Sensações não podem ser compartilhadas, no sentido estrito da palavra.

O ponto chave da discussão de Gibson, que eventualmente é o mais importante para este trabalho, é que essa definição de ambientes individual e coletivo se reflete em uma estrutura ecológica que Gibson chamou de *Arranjo Ótico do Ambiente* ou AOA (GIBSON, 1979). O ambiente individual se reflete em um *Arranjo Ótico* único para cada observador; e a locomoção constante desse observador muda constantemente esse arranjo. A figura 8 é uma representação de como se comporta o arranjo ótico de ambiente para um observador. Há um AOA unicamente definido para cada ponto do ambiente e cada um deles provê informações únicas do ambiente. Esses AOAs definidos são estáticos, enquanto houver persistência do meio. Fisicamente falando, eles se configuram como o padrão de luz converge para o ponto em questão no espaço. Quando o indivíduo se locomove, altera-se o padrão de luz convergente e adicionam-se informações temporais do ambiente, gerando um AOA dinâmico. Esse AOA pode ser representado pelo conceito de Fluxo Ótico apresentado no capítulo 3. É através desse arranjo dinâmico que o observador percebe os eventos ecológicos ao seu redor. Ele contém em si, portanto, todas as informações necessárias do meio (GIBSON, 1966). É essa afirmação que motiva essa pesquisa. Utilizando as técnicas de representação de Fluxo Ótico, a princípio, é possível reproduzir todo o arranjo ótico dinâmico e traduzir as informações ecológicas sobre o ambiente.

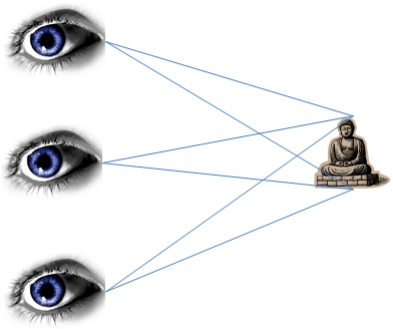

Figura 8: Representações do Arranjo Ótico Dinâmico do Ambiente
A ideia do AOA é que diferentes caminhos de visão oferecem diferentes informações sobre o ambiente.
Fonte: Autoria própria.

3 MOVIMENTO E TÉCNICAS DE CÁLCULO DE FLUXO ÓTICO

Movimento é um fenômeno físico, no qual um objeto varia a sua posição espacial ao longo do tempo. Para os seres vivos, de modo geral, a capacidade de perceber o movimento ao redor e reagir é essencial à própria sobrevivência. Evolutivamente, o Sistema Visual Humano tornou-se mais sensível à percepção de movimentos do que de imagens estáticas (GIBSON, 1979; PALMER, 1999) e é dada essa importância que se alavancou o desenvolvimento extensivo da área de Visão Computacional.

Quando falamos em percepção de movimento, falamos da capacidade dos indivíduos de identificar o movimento dos objetos ao seu redor e inferir informações como trajetória e posição espacial. O capítulo 2 mostrou que não há consenso sobre como a percepção de movimento ocorre em âmbito psicológico. Entretanto, em nível fisiológico, podemos encarar o movimento como a variação no tempo do padrão de luz que chega na retina. É a partir dessa definição que se moldam os conceitos de Visão Computacional que serão abordados ao longo deste capítulo.

Palmer (1999) considera que o movimento, no que se refere ao Sistema Visual Humano, precisa ser discriminado em dois conceitos distintos e necessários para o entendimento dos problemas de movimento computacional: movimento de objetos e movimento de imagens. A movimentação de objetos é a que acontece no mundo real tridimensional e é o objeto final que o estudo de Visão Computacional quer identificar; a movimentação de imagens é a projeção da movimentação de objetos no plano da retina. Constitui-se de um movimento em uma projeção bidimensional e é o ponto de partida do estudo de VC. O principal problema da estimação computacional de movimento consiste em modelar o movimento real de um objeto dadas as informações adquiridas da movimentação das imagens captadas, por exemplo, por uma câmera.

Para tanto, parte-se da ideia apresentada no capítulo 2 sobre o *Arranjo Ótico Dinamico de Ambiente* de Gibson que, em outros termos, é uma estrutura vetorial em que todos os vetores convergem para um ponto estacionário. Esses vetores conteriam toda a informação necessária para a determinação do movimento dos objetos no mundo real, que é o objetivo final

da percepção de movimento (GIBSON, 1979). Entretanto, essa ideia permanecia em âmbito da psicologia, com uma vaga definição geométrica. O trabalho posterior de Longuet-Higgins e Prazdny (1980) forneceu um primeiro arcabouço matemático para as ideias de Gibson, através da formulação de um modelo matemático para representar o movimento e o campo de Fluxo Ótico através de um ambiente estático dadas transições de imagens na retina quando as velocidades angulares e de translação são conhecidas. Este assunto será abordado mais profundamenteao longo deste capítulo.

Quando os computadores pessoais começaram a entrar com mais facilidade no meio acadêmico, por volta da década de 1980, surgiram os primeiros algoritmos puramente computacionais de processamento de vídeo e de estimação de movimento. Baseados nos trabalhos de Moravec (1980), Lucas e Kanade apresentaram um algoritmo pioneiro para a estimação de movimento em vídeos (LUCAS; KANADE, 1981), assim como Horn e Schunck (1980) e posteriormente Farnebäck (2002).

Entretanto, não se fala mais em Sistema Visual Humano, informações ecológicas e teoria de percepção. Fala-se agora de Câmera, Sistema de Captação de imagens, Quadros, Pixels, variações no gradiente de intensidade e softwares de processamento, no melhor estilo da doutrina “Gestalt”.

Podemos agora considerar o Campo de Fluxo Ótico como um campo vetorial, no qual cada vetor possui amplitude e direção próprias e corresponde ao deslocamento estimado de um pixel (ou um conjunto de pixels) entre dois quadros consecutivos de um vídeo. Da mesma forma, podemos considerar um vídeo como um encadeamento temporal ordenado de imagens altamente correlacionadas entre si.

Este capítulo vai apresentar técnicas de estimação de movimento (Fluxo Ótico) utilizadas nos experimentos e para tanto, a seção 3.2 vai apresentar o algoritmo de Lucas e Kanade, o princípio da deteção de pontos de interesse através do algoritmo de Harris e a definição de Fluxo Ótico esparso, com exemplos; a seção 3.3 vai apresentar o algoritmo de Horn e Schunck (1980) e a definição de Fluxo Ótico denso, com exemplos; a seção 3.4 apresentará o trabalho de Farnebäck (2002), de que forma ele agrupou em um mesmo arcabouço elementos de várias teorias de estimação de movimento e o seu algoritmo para cálculo de Fluxo Ótico denso, com exemplo.

A seção 3.5 irá aprofundar o arcabouço matemático para a estimação parametrizada de movimento.

3.1 FLUXO ÓTICO

A literatura apresenta três abordagens distintas para a estimação de Fluxo Ótico: Diferencial, por Correlação e por Energia. Os algoritmos mais conhecidos apresentados são de abordagem Diferencial. Portanto, definir essa abordagem é essencial para o desenvolvimento deste capítulo.

Vamos considerar uma imagem I na qual cada *pixel* $\mathbf{x}$ com posição (x,y) no instante t apresenta intensidade $I(x,y,t)$. Ao assumirmos que entre dois quadros consecutivos de um vídeo o deslocamento de um pixel é muito pequeno, pode-se assumir que não há variação no valor da sua intensidade, portanto:

$$I(x,y,t) = I(x+\delta x, y+\delta y, t+\delta t) \tag{1}$$

Quando assumimos deslocamentos pequenos, ou seja, pequenos valores para δx, δy e δt, possibilitamos uma expansão de Taylor de primeira ordem:

$$I(x,y,t) = I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t \tag{2}$$

que pode ser simplificada como:

$$\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = \frac{\partial I}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial I}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial I}{\partial t}\frac{\delta t}{\delta t} = 0 \tag{3}$$

O segundo termo da equação 3 é particularmente interessante, pois pode ser representado explicitamente pela velocidade do pixel em x e em y

$$v = \left(\frac{\delta x}{\delta t}, \frac{\delta y}{\delta t}\right) \tag{4}$$

e do gradiente de intensidades

$$\nabla I = \frac{\partial I}{\partial x} + \frac{\partial I}{\partial y} \tag{5}$$

que constitui a equação de restrição de Fluxo Ótico

$$\nabla I \cdot \mathbf{v} + \frac{\partial I}{\partial t} = 0 \tag{6}$$

A equação (6) por si só não é capaz de definir o Fluxo Ótico (encontrar o vetor **v**), pois são duas as variáveis e apenas uma equação, resultando no conhecido "problema de abertura" (ULLMAN, 1979). A equação de Fluxo Ótico é um problema "mal posto", pois aceita um número infinito de soluções. A partir deste ponto, são necessárias outras restrições para se achar o valor de **v** e é aí que os diversos algoritmos de estimação de Fluxo Ótico diferem entre si.

3.2 O ALGORITMO DE LUCAS E KANADE E O FLUXO ÓTICO ESPARSO

A maior contribuição do algoritmo proposto por Lucas e Kanade (1981) é a utilização de uma técnica de registro - ou alinhamento - de imagens, na qual partes de uma imagem são trasladadas, movidas e eventualmente deformadas de forma a maximizar o casamento com os pixels de uma outra imagem. De acordo com Baker e Matthews (2004), essa é a técnica mais utilizada nos algoritmos de visão computacional para registro de imagens.

O problema a ser resolvido era justamente estimar movimento entre duas imagens consecutivas de um vídeo, da forma menos custosa possível computacionalmente. Os autores perceberam que na maioria dos casos, as partes de uma imagem estão espacialmente próximas entre dois quadros consecutivos. Assim, muito menos interações seriam necessárias para se achar a região de melhor casamento.

Considere uma região dentro de uma imagem, descrita por uma função $F(x)$, que representa o valor do pixel dada sua posição. Se em uma segunda imagem, a mesma região é agora representada por uma outra função, digamos $G(x)$, então existe um valor h tal que

$$F(h+x) = G(x) \tag{7}$$

Logo, para o caso de translação pura, o problema resume-se a achar um valor de h para o qual a relação seja válida. Entretanto, como em aplicações reais existe a presença de ruído e é pouco usual um caso de translação pura, o problema torna-se:

$$\sum_{x \in \mathbb{R}} |F(h+x) - G(x)|^2 = \varepsilon \tag{8}$$

onde ε é o somatório dos erros de estimação. Nesse caso, deve-se achar um valor de h tal que o erro ε seja minimizado, tornando-se um problema de otimização.

A principal crítica dos autores quanto aos possíveis métodos de otimização da época é

que eles são ou muito custosos computacionalmente e demorados ou passíveis de falha na busca pelo valor ótimo de h. Para contornar esse problema, eles propuseram um algoritmo de aprendizado similar ao de Newton-Raphson. Inicialmente, arbitra-se um valor para h e o gradiente de intensidade em cada *pixel* da imagem é usado para atualizá-lo até que se atinja a convergência. Dessa forma, para uma imagem com tamanho NxN e uma região de tamanho MxM, diminui-se a complexidade computacional de $O(M^2N^2)$ (busca exaustiva) para $O(M^2logN)$, dependendo da estimação do valor inicial de h (LUCAS; KANADE, 1981).

De uma forma geral para um cenário n-dimensional, o objetivo é minimizar o erro na equação 8 em relação a h. Isso implica a condição

$$\frac{\partial \varepsilon}{\partial h} = 0 \tag{9}$$

Se considerarmos h um valor pequeno, pode-se aproximar a derivada parcial de $F(x)$ para:

$$\frac{\partial F(x)}{\partial x} \approx \frac{F(x+h) - F(x)}{h} \tag{10}$$

A aproximação em (10) pode ser usada no problema de minimização em (9), de forma que

$$\begin{aligned} \frac{\partial}{\partial h} \sum_R \left[F(x) + h\frac{\partial F(x)}{\partial x} - G(x) \right]^2 = 0 \\ \sum_R 2\frac{\partial F(x)}{\partial x} \left[F(x) + h\frac{\partial F(x)}{\partial x} - G(x) \right] = 0 \end{aligned} \tag{11}$$

Assim, tira-se que

$$h = \left[\sum_R \left(\frac{\partial F(x)}{\partial x} \right)^T [G(x) - F(x)] \right] \left[\sum_R \left(\frac{\partial F(x)}{\partial x} \right)^T \left(\frac{\partial F(x)}{\partial x} \right) \right]^{-1} \tag{13}$$

A partir dessa estimação, inicia-se um sistema iterativo, movendo-se repetidamente $F(x)$ de uma distância estimada h e estimando-se um novo h até alcançar a convergência. Logo, podemos resumir o algoritmo de Lucas-Kanade como um processo iterativo de acordo com o algoritmo 1.

Algoritmo 1 Algoritmo de Lucas-Kanade

$h_0 \leftarrow 0$

while $\Delta h \neq 0$ **do**

$$h_{k+1} \leftarrow h_k + \frac{\left[\sum_R \left(\frac{\partial F(x+h_k)}{\partial x}\right)^T [G(x) - F(x+h_k)]\right]}{\left[\sum_R \left(\frac{\partial F(x+h_k)}{\partial x}\right)^T \left(\frac{\partial F(x+h_k)}{\partial x}\right)\right]}$$

$\Delta h \leftarrow h_{k+1} - h_h$

end while

A figura 9 mostra um campo de Fluxo Ótico esparso calculado pelo algoritmo de Lucas-Kanade. As setas indicam a direção e a intensidade dos vetores de Fluxo Ótico estimados. Pode-se observar que os Pontos de Interesse utilizados para a estimação encontram-se em regiões específicas da imagem, nas quais a informação visual é mais relevante. Convém observar que a presença de regiões de alta frequência na imagem geram ruídos e erros de estimação.

3.2.1 A DETEÇÃO DE PONTOS DE INTERESSE

O algoritmo de Lucas-Kanade diminui consideravelmente o tempo de processamento quando comparado a uma busca exaustiva. Entretanto, uma questão que não foi considerada pelos autores é a de estimação de movimento. Originalmente, esse algoritmo era para alinhamento de imagens, mas poderia ser estendido para o caso de Fluxo Ótico.

Surge uma questão de caráter prático que precisa ser respondida: como, quantas e quais regiões de interesse devem ser utilizadas para estimação de movimento em um vídeo? O caráter baseado em erro por alinhamento de imagens de Lucas-Kanade possui uma limitação muito grande em imagens com superfícies lisas e homogêneas. Nesse caso, a condição na equação (9) seria válida para mais de um valor de h. Esse é um grande indicativo de que deve existir um conjunto de regiões ótimas para as quais esse algoritmo funcionaria bem em estimação de movimento.

Moravec (1980) apresentou um primeiro algoritmo para navegação autônoma de robôs - mais precisamente para exploração do solo marciano - cujo objetivo era evitar obstáculos através da análise de imagens. Após calibração da câmera, um procedimento, chamado pelo autor de *operador de interesse*, selecionava 30 ou mais pontos distintos na imagem, pontos de interesse ou PIs, com características únicas mais marcantes entre eles. Segundo Moravec, uma *característica* é relevante se ela pode ser facilmente localizada sem ambiguidades entre diferentes pontos de vista em uma cena.

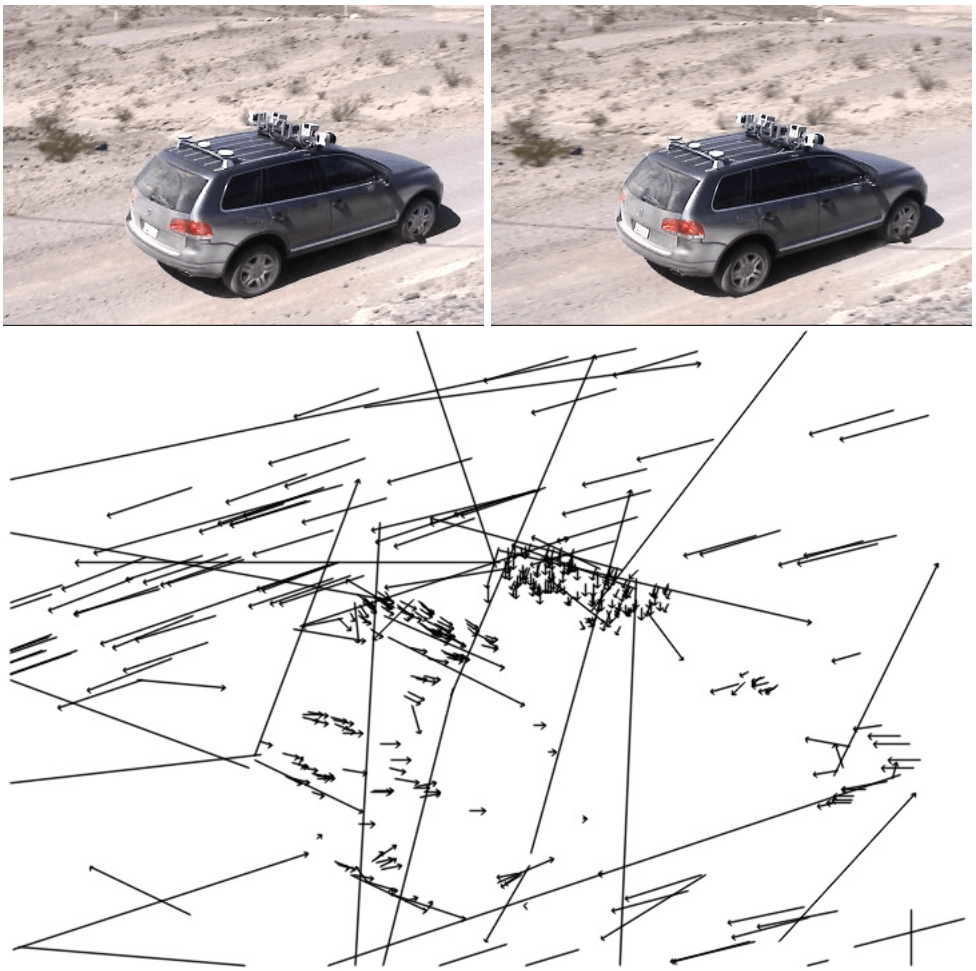

Figura 9: Fluxo Ótico estimado por Lucas e Kanade.
Exemplo de estimação de um vetor de Fluxo Ótico esparso por Lucas e Kanade. As duas primeiras imagens representam dois quadros consecutivos do vídeo de controle. Cada vetor do campo de Fluxo Ótico estimado (imagem maior) representa o deslocamento de um ponto de interesse entre os dois quadros, uma vez que o campo estimado é esparso.
Fonte: Autoria própria.

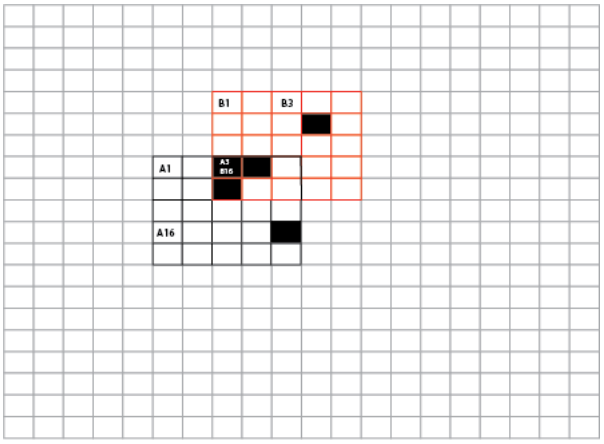

Figura 10: Máscaras para soma dos quadrados, de acordo com Moravec.
Fonte: Adaptação de Moravec (1980).

Na prática, a primeira medida de quanto um ponto é distinto é a variância direcional sobre uma janela quadrada de pixels. Cada uma dessas janelas é deslocada em todas as direções e a média dos valores de intensidade são computadas. Quando se encontra um máximo local ou global, um ponto de interesse é definido. Deslocamentos sobre uma superfície homogênea geram uma pequena - ou nenhuma - variação na medida de intensidade. Em bordas, a variação na medida de intensidade é pequena para deslocamentos na direção das bordas. Dessa forma, o que Moravec sugeriu é um detetor de cantos, pois grandes mudanças - os máximos locais ou globais - ocorrem em cantos.

A figura 10 mostra uma janela de pixels deslocada com seus respectivos valores de intensidade. Podemos resumir o comportamento do detedor de Moravec através da equação 14:

$$E_{x,y} = \sum_{u,v} w_{u,v} |I_{x+u,y+v} - I_{u,v}|^2 \tag{14}$$

onde $I_{u,v}$ é a intensidade do pixel na posição (u,v), $E_{x,y}$ é a mudança produzida por um deslocamento (x,y) e $w_{u,v}$ é um coeficiente de amortecimento, normalmente de valor unitário ou uma janela gaussiana.

Contanto, esse detetor possui limitações, por exemplo (HARRIS; STEPHENS, 1988):

- A resposta é anisotrópica
- A resposta está sujeita à interferência de ruído
- Somente o valor de E é considerado.

Para contornar essas limitações, Harris e Stephens (1988) sugerem algumas medidas corretivas.

A resposta do detetor de Moravec é anisotrópica porque considera só deslocamentos nas 8 direções cardeais principais. A equação de Moravec (14) pode ser expandida ao redor do deslocamento de origem, para considerar todos os possíveis deslocamentos:

$$E_{x,y} = \sum_{u,v} w_{u,v} \left[x\frac{\partial I}{\partial x} + y\frac{\partial I}{\partial y} \right] \tag{15}$$

e por sua vez,

$$E_{x,y} = Ax^2 + 2Cxy + By^2, \tag{16}$$

onde os valores A, B e C são representados pela convolução das derivadas da intensidade I pelo vetor w.

$$\begin{aligned} A &= \left(\frac{\partial I}{\partial x}\right)^2 \otimes w \\ B &= \left(\frac{\partial I}{\partial y}\right)^2 \otimes w \\ C &= \left(\frac{\partial I}{\partial x}\frac{\partial I}{\partial y}\right) \otimes w \end{aligned} \tag{17}$$

O uso de janelas retangulares e binárias torna o detetor pouco robusto a ruídos. O uso de uma janela circular, com fator de suavização seguindo uma função Gaussiana

$$w_{u,v} = e^{\frac{-(u^2+v^2)}{2\sigma^2}} \tag{19}$$

é sugerido como a forma mais adequada (HARRIS; STEPHENS, 1988).

Por fim, os autores sugerem que a equação (16) seja reescrita de forma matricial, isolando as constances A, B e C das variáveis x e y, da forma

$$E_{x,y} = (x,y)\mathbf{M}(x,y)^T \tag{20}$$

onde $\mathbf{M}$ é uma matriz 2×2 que descreve a função de autocorrelação na origem (sem desloca-

mento).

$$\mathbf{M} = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \tag{21}$$

A matriz $\mathbf{M}$ possui dois autovalores α e β. Existem 3 casos que os autores consideram, para os valores de α e β. Se ambos forem pequenos, isso indica que as mudanças E de intensidade sobre a janela $w_{u,v}$ são pequenas para cada deslocamento possível; logo, trata-se de uma região homogênea. Se ambos os valores forem grandes, os valores das mudanças em E são grandes em qualquer direção de deslocamento, portanto, possivelmente trata-se de um canto. Já se o valor de α é muito maior que o de β ou o contrário, isso quer dizer que o valor da mudança em E é grande apenas em uma direção, sendo um indicativo de bordas.

Uma forma de refinar a classificação dos pixels em bordas e cantos é a formulação de uma medida de qualidade da classificação. Os autores sugerem uma função de resposta R:

$$R = \alpha + \beta - k(\alpha\beta)^2 \tag{22}$$

ou, para os elementos da matriz $\mathbf{M}$:

$$R = A + B - k(AB - C^2)^2 \tag{23}$$

Com esse indicador, pode-se dizer que valores positivos de R indicam cantos; valores negativos indicam bordas; e valores próximos a zero indicam superfície homogênea. Por questões práticas, aplicam-se dois limiares para garantir a continuidade das bordas e refinar a classificação.

A figura 11 é um exemplo da utilização do detetor de Harris e Stephens. Os círculos na imagem correspondem aos pontos com máximos valores de R, ou seja, cantos. Existe uma grande vantagem prática na utilização desses pontos para a estimação de Fluxo Ótico: parte-se de pressuposto que duas imagens consecutivas de um vídeo são muito parecidas, logo, são esses pontos, considerados mais singulares, que espera-se encontrar em ambas.

Shi e Tomasi (1994) abordaram o problema da deteção de pontos de interesse pontualmente para o caso de quadros consecutivos de um vídeo. Eles observaram que muitos dos pontos mais distintos de uma imagem, tais como cantos e bordas, eventualmente sofrem oclusão no quadro seguinte de um vídeo. O problema de rastreamento de pontos em um vídeo continua um problema mal enunciado devido a essa condição.

Figura 11: Exemplo de deteção de pontos de interesse por Harris.

Fonte: Autoria própria.

Os autores sugerem uma medida da qualidade de um ponto para se tornar ponto de interesse baseado não somente nas informações intra-quadros, mas com a facilidade de ser identificado no quadro seguinte de um vídeo. Surge, então, a medida de dissimilaridade.

Considere dois quadros consecutivos de um vídeo. Dissimilaridade é a medida da diferença de assinatura entre um ponto de interesse do primeiro quadro e de um ponto de interesse no segundo quadro. Ou seja, altos valores de dissimilaridade implicam que o ponto não é o ideal para ser referido como ponto de interesse.

A hipótese inicial e necessária para essa medida é a de que a posição de um ponto de interesse com a menor dissimilaridade no segundo quadro é apenas ligeiramente diferente da posição do ponto de interesse no primeiro quadro. Seja, portanto, a equação que relaciona as intensidades dos pixels entre os dois quadros

$$I_2(x,y,t+\tau) = I_1(x-\xi(x,y,t,\tau), y-\eta(x,y,t,\tau)) \tag{24}$$

e seja

$$\delta = (\xi, \eta) \tag{25}$$

o deslocamento de um ponto em uma posição $\mathbf{x} = (x,y)$.

Essa equação é uma releitura da equação 6. Podemos reescrevê-la sob a representação de um movimento afim:

$$\delta = D\mathbf{x} + \mathbf{d} \tag{26}$$

onde D é a chamada *matriz de deformação*

$$D = \begin{bmatrix} d_{xx} & d_{xy} \\ d_{yx} & d_{yy} \end{bmatrix} \tag{27}$$

e $\mathbf{d}$ é o vetor de translação.

Então, a equação 24 pode ser reescrita em termos de D, $\mathbf{x}$ e $\mathbf{d}$ como

$$I_2[(\mathbf{1}+D)\mathbf{x}+\mathbf{d}] = I_1(\mathbf{x}) \tag{28}$$

Existe uma restrição muito grande quando se fala dessa forma de rastreamento. Ao redor de um ponto de interesse existe uma janela de pixels cuja variação da intensidade é sua assinatura. Essa janela é utilizada para se estimar os valores da matriz de deformação D e do vetor de translação $\mathbf{d}$. Quanto menor a janela, torna-se mais difícil estimar tais valores, porém, menos suscetível a efeitos de profundidade.

Dado que a quantidade de ruído dentro de uma imagem pode ser grande, a equação (28) pode não ser satisfeita. A sugestão dos autores para uma melhor estimativa da dissimilaridade é a equação:

$$\varepsilon = \int\int_W [I_2[(\mathbf{1}-D)\mathbf{x}+\mathbf{d}]-I_1(\mathbf{x})]^2 w(\mathbf{x})d\mathbf{x} \tag{29}$$

onde W é a janela ao redor do ponto e $w(\mathbf{x})$ é uma função de peso qualquer, normalmente uma gaussiana. O que resume o problema a encontrar os valores de D e $\mathbf{d}$ que minimizam a dissimilaridade ε. A figura 12 é um exemplo da deteção de pontos de interesse via dissimilaridade. Em comparação com o método de Harris e Stephens, os Pontos de Interesse de Shi e Tomasi são mais estáveis ao longo de um conjunto de imagens e o seu custo computacional é menor, devido às suposições feitas no processo de cálculo das dissimilaridades. Os resultados apresentados nas figuras 9 e 12 são coerentes entre si, visto que estimam os mesmos pontos. Entretanto, a estimação por Shi e Tomasi é menos restritiva, considerando mais pontos de interesse que Harris e Stephens. Ao longo deste trabalho, o método de Shi e Tomasi mostrou-se mais adequado, pelo ganho no tempo de processamento dos algoritmos apresentados no capítulo 5.

3.3 O ALGORITMO DE HORN E SCHUNCK E O FLUXO ÓTICO DENSO

A equação de restrição do Fluxo Ótico (6) pode ser reescrita como

$$\nabla I \cdot \mathbf{v} = -I_t \tag{30}$$

Escrita dessa forma, a equação de Fluxo Ótico é capaz de calcular a componente de movimento na direção do gradiente de intensidade

$$|v_{//}| = -\frac{I_t}{\sqrt{I_x^2+I_y^2}} \tag{31}$$

Entretanto, ainda é necessário calcular uma segunda componente de movimento, a fim

Figura 12: Exemplo de deteção de pontos de interesse por Shi e Tomasi.

Fonte: Autoria própria.

de se conseguir ambas as componentes cartesianas de **v**. Portanto, alguma outra restrição deve ser imposta.

Os autores, nesse ponto (e essa é a principal diferença entre o algoritmo de Horn e Schunck e o de Lucas e Kanade), sugerem que se não houver oclusão de objetos, as velocidades e a intensidade dos pixels em uma pequena vizinhança variam muito pouco. Assim, as informações das derivadas de maior ordem das componentes de velocidade x e y seriam sempre zero - ou o menor valor possível. Isso dá um bom indício para o uso do operador laplaciano sobre o vetor de velocidade $\mathbf{v} = (v_x, v_y)$. Sejam os operadores Laplaciano para as componentes de **v**, conforme as equações a seguir.

$$\begin{aligned} \nabla^2 v_x &= \frac{\partial^2 v_x}{\partial x^2} + \frac{\partial^2 v_x}{\partial y^2} \\ \nabla^2 v_y &= \frac{\partial^2 v_y}{\partial x^2} + \frac{\partial^2 v_y}{\partial y^2} \end{aligned} \tag{32}$$

Enquanto na definição de Fluxo Ótico esparso, a escolha de pontos com características peculiares de variação de intensidades era o tema principal, a hipótese da variação suave em uma vizinhança dá ao algoritmo de Horn e Schunck a conveniência de fugir da discussão sobre pontos de interesse e escolher pontos espaçados igualmente ao longo de toda a imagem, ou seja, cuja distância em *pixels* de um ponto ao outro seja constante. Cria-se, então um campo de pontos ou uma grade sobre a imagem, caracterizando o Fluxo Ótico obtido como *denso*.

Computacionalmente, a desvantagem da abordagem densa de Horn e Schunck é que as derivadas parciais da intensidade I em relação a t, x e y e os Laplacianos de v_x e v_y precisam ser calculados para um grande número de pontos. Os autores sugerem que a característica quantizada da intensidade em uma imagem permite estimações razoáveis de tais grandezas. Para as estimações das derivadas de intensidade, os autores apresentam uma abordagem que se baseia em um cubo (x, y, t) de tamanho $2x2x2$. A relação das intensidades em cada vértice do cubo dá as estimações das derivadas da intensidade em cada uma das direções. Seja um pixel (i, j, t). Considerando o mesmo pixel entre dois quadros consecutivos de um vídeo, forma-se o cubo representado na figura 13.

As derivadas de intensidade nesse ponto são definidas como:

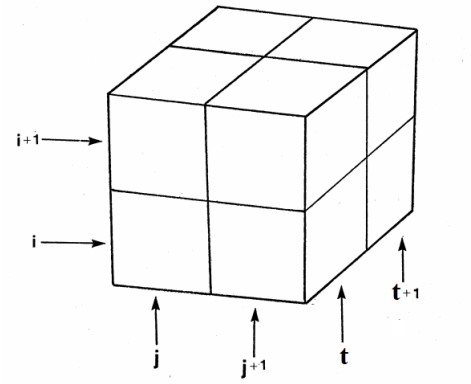

Figura 13: Máscara de aproximação de derivadas parciais.
Cubo que representa a aproximação das derivadas parciais de intensidade.
Fonte: Horn e Schunck (1980)

$$I_x \approx \frac{1}{4}(I_{i,j+1,t} - I_{i,j,t} + I_{i+1,j+1,t} - I_{i+1,j,t} + I_{i+1,j,t+1} - I_{i,j,t+1} + I_{i+1,j+1,t+1} - I_{i+1,j,t+1}) \quad (34)$$
$$I_y \approx \frac{1}{4}(I_{i+1,j,t} - I_{i,j,t} + I_{i+1,j+1,t} - I_{i,j+1,t} + I_{i+1,j,t+1} - I_{i,j,t+1} + I_{i+1,j+1,t+1} - I_{i,j+1,t+1})$$
$$I_t \approx \frac{1}{4}(I_{i,j,t+1} - I_{i,j,t} + I_{i+1,j,t+1} - I_{i+1,j,t} + I_{i+1,j,t+1} - I_{i,j+1,t} + I_{i+1,j+1,t+1} - I_{i+1,j+1,t})$$

Já os Laplacianos de v_x e v_y são definidos em termos das suas médias ($\hat{v}_x$ e $\hat{v}_y$), também estimadas:

$$\nabla^2 v_x \approx \kappa(\hat{v}_{x,i,j,k} - v_{x,i,j,k}) \quad (36)$$
$$\nabla^2 v_y \approx \kappa(\hat{v}_{y,i,j,k} - v_{y,i,j,k})$$

onde os valores de $\hat{v}_x$ e $\hat{v}_y$ são aproximados por

$$\hat{v}_{x,i,j,t} \approx \frac{1}{6}v_{x,i-1,j,t} + v_{x,i,j+1,t} + v_{x+1,i,j,t} + v_{x,i,j-1,t} \quad (38)$$
$$+\frac{1}{12}v_{x,i-1,j-1,t} + v_{x,i-1,j+1,t} + v_{x,i+1,j+1,t} + v_{x,i+1,j-1,t}$$
$$\hat{v}_{y,i,j,t} \approx \frac{1}{6}v_{y,i-1,j,t} + v_{y,i,j+1,t} + v_{y+1,i,j,t} + v_{y,i,j-1,t} \quad (39)$$
$$+\frac{1}{12}v_{y,i-1,j-1,t} + v_{y,i-1,j+1,t} + v_{y,i+1,j+1,t} + v_{y,i+1,j-1,t}$$

O valor de κ é, segundo os autores, igual a 3 para essa aproximação das médias. Essa

abordagem representa a aplicação de uma máscara $3x3$ em uma vizinhança do ponto considerado.

A utilização de estimações que se baseiam em outras estimações aumentam a magnitude do erro encontrado. O problema inicial era minimizar a soma dos erros da estimação da taxa de variação das intensidades dos pixels ao longo da imagem:

$$\varepsilon_b = I_x v_x + I_y v_v + I_t \tag{41}$$

onde ε_b é o erro de estimação da intensidade dos pixels.

Entretanto, a introdução de mais erros de estimação precisa ser compensada. Para isso, os autores sugerem um segundo problema a ser resolvido, que é minimizar o erro de estimação das velocidades, ou seja:

$$\alpha^2 \varepsilon_c^2 = (\hat{v}_x - v_x)^2 + (\hat{v}_y - v_y)^2 \tag{42}$$

onde α é um fator de peso que deve ser convenientemente pequeno (HORN; SCHUNCK, 1980) e ε_c é o erro de estimação das velocidades.

Com essas duas medidas, pode-se traduzir o problema como encontrar os valores de v_x e v_y que minimizam um erro ε definido por

$$\varepsilon^2 = (\alpha \varepsilon_c)^2 + \varepsilon_b^2 \tag{43}$$

É de se esperar, dadas as suposições de pequenas variações na intensidade dentro de uma pequena vizinhança, que as derivadas de maior ordem (em relação a v_x e v_y) desses erros sejam zero. Essa condição nos dá uma restrição nos valores de v_x e v_y.

$$\begin{aligned} \frac{\partial \varepsilon^2}{\partial v_x} = 0 &= -2\alpha^2(\hat{v}_x - v_x) + 2(I_x v_x + I_y v_y + I_t) I_x \\ \frac{\partial \varepsilon^2}{\partial v_y} = 0 &= -2\alpha^2(\hat{v}_y - v_y) + 2(I_x v_x + I_y v_y + I_t) I_y \end{aligned} \tag{44}$$

A manipulação algébrica da equação (44) nos leva a um par de equações a serem resolvidas para cada ponto da grade sobre a imagem:

$$v_x = \hat{v}_x - I_x \frac{I_x \hat{v}_x + I_y \hat{v}_y + I_t}{\alpha^2} \qquad (46)$$
$$v_y = \hat{v}_y - I_y \frac{I_x \hat{v}_x + I_y \hat{v}_y + I_t}{\alpha^2}$$

É importante observar que na época em que tal abordagem foi proposta, era impraticável resolver ambas as equações por meios tradicionais. Mesmo nos computadores modernos, a quantidade de equações que precisam ser resolvidas simultaneamente para todas as imagens da grade é muito grande. A saída encontrada foi a utilização de um algoritmo iterativo em que as estimativas das velocidades atuais dependem unicamente das estimativas das derivadas parciais de intensidade e das estimativas das velocidades médias anteriores, portanto:

$$v_x^{n+1} = \hat{v}_x^n - I_x \frac{I_x \hat{v}_x^n + I_y \hat{v}_y^n + I_t}{\alpha^2} v_y^{n+1} = \hat{v}_y^n - I_y \frac{I_x \hat{v}_x^n + I_y \hat{v}_y^n + I_t}{\alpha^2} \qquad (48)$$

Assim, podemos resumir o método de Horn e Schunck através do algoritmo 2.

Algoritmo 2 Algoritmo de Horn e Schunck

$\mathbf{v}_0 \leftarrow \mathbf{0}$
while $\mathbf{v}^{n+1} - \mathbf{v}^n > \varepsilon$ **do**
$v_x^{n+1} \leftarrow \hat{v}_x^n - I_x \frac{I_x \hat{v}_x^n + I_y \hat{v}_y^n + I_t}{\alpha^2}$
$v_y^{n+1} \leftarrow \hat{v}_y^n - I_y \frac{I_x \hat{v}_x^n + I_y \hat{v}_y^n + I_t}{\alpha^2}$
end while

A figura 14 mostra um exemplo de estimação de Fluxo Ótico pela abordagem densa de Horn e Schunck. Os vetores de Fluxo Ótico são calculado baseados em uma vizinhança de *pixels*, mostram a direção e intensidade de movimentação de blocos de *pixels* e se espalham por uma grade regular, neste caso, um quadrado de 16 *pixels* de lado. O procedimento de Horn e Schunck é computacionalmente pesado, pois é aplicado à cada vizinhança ao longo de todos os pares de imagens consecutivas ao longo do vídeo.

3.4 O ALGORITMO DE FARNEBÄCK

O trabalho inicial de Farnebäck não era *a priori* uma forma alternativa de estimação de vetores de Fluxo Ótico. Ele observou que muitos dos algoritmos presentes na literatura recaíam sobre uma figura matemática chamada *Tensor Estrutural* e que esse era um indício de que deveria haver um arcabouço matemático unificado para a estimação de movimento.

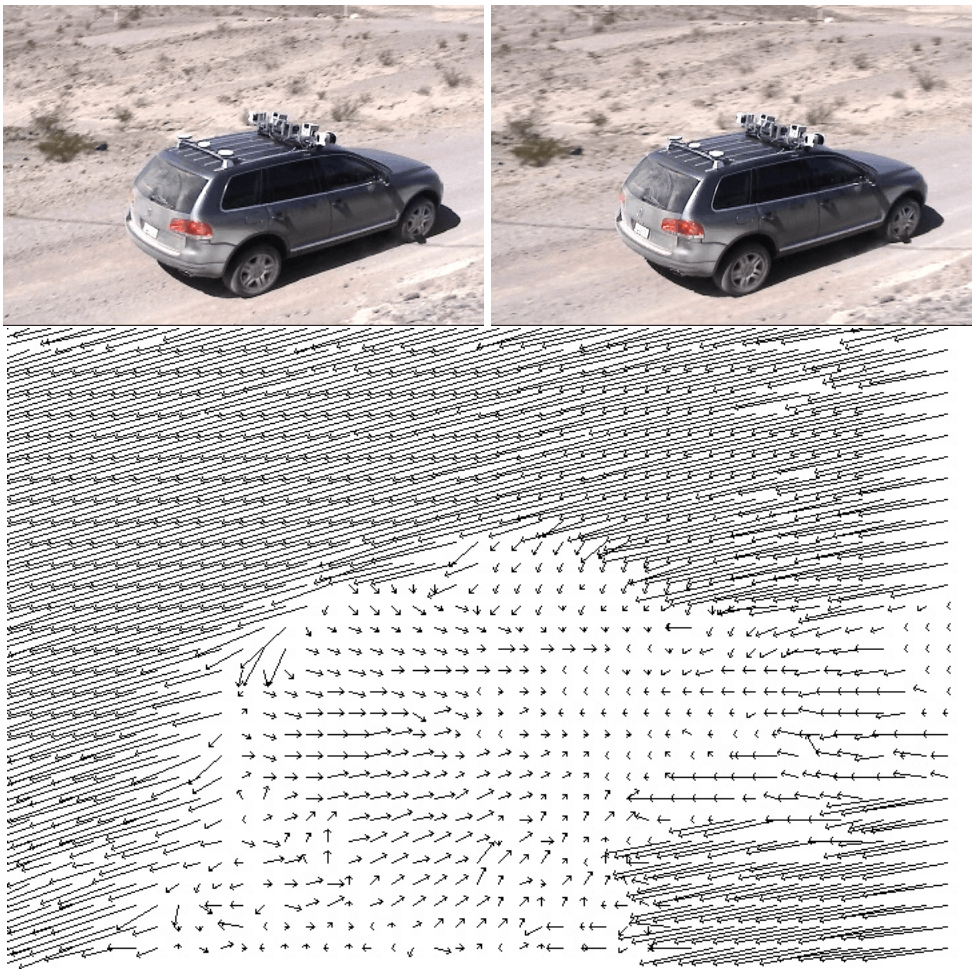

Figura 14: As duas primeiras imagens representam dois quadros consecutivos do vídeo de controle. A imagem maior é o campo de Fluxo Ótico denso estimado para os dois quadros usando blocos de 16 *pixels*.

Fonte: Autoria própria.

A partir da definição da equação de restrição de velocidade de Fluxo Ótico (6) pode-se chegar em uma segunda definição, através de uma modificação algébrica:

$$(\nabla I)^T(\nabla I)\mathbf{v} = \mathbf{0} \tag{50}$$

O que Farnebäck notou é que essa segunda definição implica que um autovetor de $(\nabla I)^T(\nabla I)$ cujo autovalor seja 0 corresponderia a $\mathbf{v}$. Apesar de essa informação não ser por si só suficiente para estimar o valor de $\mathbf{v}$, se o valor médio do produto dos gradientes puder ser estimado em uma região na qual a velocidade $\mathbf{v}$ possa ser considerada constante, então ela pode ser estimada. Dessa forma, a equação (50) torna-se

$$[\int_{\Omega} p(\mathbf{x})(\nabla I_{\mathbf{x}})^T(\nabla I_{\mathbf{x}})d\mathbf{x}]\mathbf{v} = \mathbf{0} \tag{51}$$

onde $p(\mathbf{x})$ é uma função de amortecimento, comumente gaussiana.

A integral do lado esquerdo da fórmula (51) é um *Tensor Estrutural* e é a partir dessa entidade matemática que o trabalho de Farnebäck se desenvolve.

Para entendê-lo, primeiramente precisamos formalizar um vídeo como um um volume espaço-temporal, no qual duas coordenadas são os eixos x e y e a terceira coordenada é o tempo t. O Tensor Estrutural carrega em si informações de orientação e velocidade da configuração tridimensional do vídeo, em termos de seus autovalores e autovetores. Isso justifica o uso desse tensor como forma de estimação de movimento.

É possível perceber a relação entre o movimento das imagens no plano (x,y) e o correspondente movimento no volume espaço-temporal. Um ponto que se move no plano corresponde a uma linha oblíqua e da direção dessa linha pode-se extrair informações do movimento real de um objeto (FARNEBÄCK, 2002); uma linha corresponde a um plano, cuja orientação dá a componente normal da velocidade real. Nesse caso, a velocidade paralela não é possível de ser obtida, dado ao problema da abertura (ULLMAN, 1979).

O primeiro passo a ser considerado é a aproximação de uma região ao redor de cada pixel através de um polinômio de segunda ordem, ou seja:

$$p(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c \tag{52}$$

onde

$$\begin{aligned} \mathbf{x} &= \begin{pmatrix} x \\ y \end{pmatrix} \qquad (53) \\ \mathbf{A} &= \begin{pmatrix} r_4 & \frac{r_6}{2} \\ \frac{r_6}{2} & r_5 \end{pmatrix} \\ \mathbf{b} &= \begin{pmatrix} r_2 \\ r_3 \end{pmatrix} \qquad (54) \\ c &= r_1 \end{aligned}$$

e $r_1, r_2, ..., r_6$ são coeficientes de expansão. Farnebäck (1996) sugere que esses coeficientes podem ser obtidos através da convolução da janela ao redor de cada pixel com as funções

$$\{1, x, y, x^2, y^2, xy\}$$

Considere que entre dois quadros, o pixel de referência, **x**, deslocou-se uma distância **d**. Dessa forma, no segundo quadro

$$\begin{aligned} p_2(\mathbf{x}) = p(\mathbf{x}-\mathbf{d}) &= (\mathbf{x}-\mathbf{d})^T \mathbf{A}(\mathbf{x}-\mathbf{d}) + \mathbf{b}^T(\mathbf{x}-\mathbf{d}) + c \qquad (56) \\ &= \mathbf{x}^T \tilde{\mathbf{A}}\mathbf{x} + \tilde{\mathbf{b}}^T\mathbf{x} + c \end{aligned}$$

com

$$\begin{aligned} \tilde{\mathbf{A}} &= \mathbf{A} \qquad (58) \\ \tilde{\mathbf{b}} &= \mathbf{b} - 2\mathbf{A}\mathbf{d} \\ \tilde{c} &= c + \mathbf{d}^T\mathbf{A}\mathbf{d} - \mathbf{b}^T\mathbf{d} \end{aligned}$$

Era de se esperar que se a expansão polinomial fosse aplicada sobre dois pontos correspondende de duas imagens, as condições em (58) valeriam. Entretanto, de acordo com o próprio autor, devido a erros de quantização e ruído, essa aproximação nem sempre é verdadeira. Portanto, como forma de reduzir possíveis erros, o valor de **A** que deve ser assumido é a média simples do valor de $\tilde{\mathbf{A}}$ calculado para o primeiro e para o segundo quadros.

As equações em (58) nos dão uma restrição de movimento a partir da qual o problema

pode ser resolvido:

$$\mathbf{A}\mathbf{d} = -\frac{1}{2}(\tilde{\mathbf{b}} - \mathbf{b}) = -\frac{\Delta\mathbf{b}}{2} \tag{60}$$

para cada pixel da imagem.

Observe que se considerarmos toda essa formulação para todos os pixels da imagem, então existe um campo de deslocamento $\mathbf{d}(x,y)$, um campo matricial $\mathbf{A}(x,y)$ e um campo de diferença $\Delta\mathbf{b}(x,y)$. É necessário assumir que o campo de deslocamento varia pouco para uma vizinhança ao redor de um pixel. Dessa forma, Ω é uma região ao redor de um pixel (x,y). O problema de estimação de movimento resume-se, agora, a achar o valor de deslocamento $\mathbf{d}(x,y)$ para o qual minimiza-se o somatório das restrições (60) ao longo de Ω:

$$e(x,y) = \sum_{\Delta x, \Delta y \in \Omega} w(\Delta x, \Delta y) \|\mathbf{A}(x+\Delta x, y+\Delta y)\mathbf{d}(x,y) - \Delta\mathbf{b}(x+\Delta x, y+\Delta y)\|^2 \tag{61}$$

O menor valor (FARNEBÄCK, 2001) seria dado por pela equação

$$\mathbf{d}(x,y) = \left(\sum w\mathbf{A}^T\mathbf{A}\right)^{-1} \sum w\mathbf{A}^T\Delta\mathbf{b} \tag{62}$$

e seria correspondente a

$$e(x,y) = \sum w\Delta\mathbf{b}^T\Delta\mathbf{b} - \mathbf{d}^T \sum w\mathbf{A}^T\Delta\mathbf{b} \tag{63}$$

Este algoritmo pode - e normalmente é - melhorado utilizando-se várias escalas diferentes. Através de uma pirâmide de escala, um processo iterativo estima o deslocamento $\mathbf{d}$ para um nível superior da pirâmide. Então, se o valor do deslocamento estiver suficientemente livre de erros, parte-se para o degrau inferior da pirâmide e usa-se o valor estimado de $\mathbf{d}$ como parâmetro de entrada. A figura 15 é um exemplo de um campo de Fluxo Ótico obtido pelo algoritmo de Farnebäck. Assim como em qualquer campo de Fluxo Ótico denso, os vetores de movimento apresentados correspondem ao deslocamento estimado de um bloco de *pixels*, nesde caso, um quadrado de lado 16. O que podemos observar subjetivamente é que os vetores de movimento de objetos, como o carro na imagem, possui um padrão ou um comportamento diferente dos vetores de fundo, associados ao *egomotion*.

O método de Farnebäck é menos custoso computacionalmente que o de Horn e Schunck,

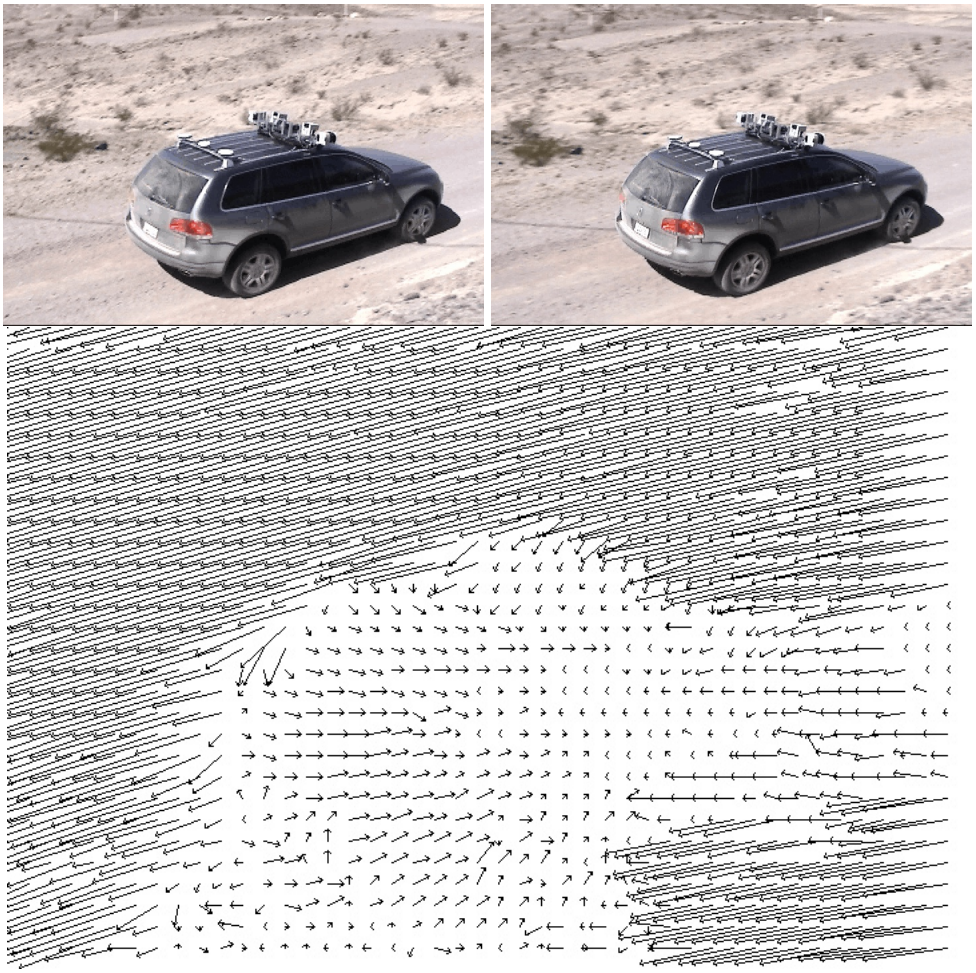

Figura 15: Exemplo de estimação de Fluxo Ótico por Farnebäck.
As duas primeiras imagens representam dois quadros consecutivos do vídeo de controle. A imagem maior é o campo de Fluxo Ótico denso estimado pelo algoritmo de Farnebäck (2002) para os dois quadros, usando blocos de 16 *pixels*.
Fonte: Autoria própria.

devido à sua característica de aproximações polinomiais. Entretanto, costuma ser mais estável e apresentando menos ruído. Esse método surgiu da tentativa de Farnebäck de unificar os métodos de estimação de Fluxo Ótico.

Existe uma diferença na execução dos algoritmos apresentados até então. Segundo Neto e Gomes (2011), a complexidade dos métodos cresce quanto maior for o tamanho da imagem em *pixels*. O algoritmo de Lucas e Kanade, por se aproximar de uma busca exaustiva, é mais custoso quanto mais pixels e pontos de interesse existem. Já Horn e Schunck e Farnebäck trabalham em uma vizinhança muito menor, o que diminui a complexidade do algoritmo e o tempo de processamento. Entretanto, os algoritmos de Fluxo Ótico denso possuem uma característica peculiar. A quantidade de pontos em uma grade de estimação pode ser maior do que a quantidade de pontos de interesse analisados em um algoritmo esparso. Quanto menor

a grade, mais tempo demora o processamento. É possível que a estimação densa associada a processamento paralelo seja a chave para o trabalho com Fluxo Ótico em tempo real.

3.5 EGOMOTION E ESTIMAÇÃO DE MOVIMENTO

A definição do arranjo ótico de Gibson, no capítulo 2, é essencial para a definição do que é um objeto independente e do que é ambiente. Se olhamos um objeto se deslocando no espaço, o deslocamento desse objeto altera o padrão luminoso que a retina recebe. Por consequência, altera o arranjo ótico e gera Fluxo Ótico. Entretanto, se um observador se desloca, é como se todo o ambiente estivesse se deslocando e modificando o arranjo ótico e igualmente gerando fluxo ótico na retina. Ao movimento do observador, dá-se o nome de *egomotion* ou movimento próprio.

Quando se fala em *egomotion*, pode-se dizer que o Fluxo Ótico gerado é relativo ao movimento entre o observador e o meio. Aqui, a definição de meio é tudo ao redor do observador que é percebido como estático. Uma bola parada sobre uma superfície faz parte do meio, enquanto uma bola rolando sobre essa superfície, não.

O Fluxo Ótico gerado pelo *egomotion* possui as mesmas características do movimento rígido, entretanto, aplicado a todo o campo visual. Vamos considerar o mecanismo geométrico da visão e dele retirar representações do movimento da imagem na retina, dadas as informações do movimento real.

A ideia de Fluxo Ótico em ambiente computacional recai sobre o conceito de que variações no padrão de captação de luz na retina criam um campo vetorial com informação sobre o ambiente e sobre o movimento em geral, conforme a teoria apresentados no capítulo 2. Este trabalho utiliza a abordagem ecológica de Gibson, segundo a qual esse campo vetorial que surge na retina consegue, por si próprio, estimar as características de movimento, dentro do mundo tridimensional.

É possível calcular geometricamente o comportamento das alterações dos padrões de luz na retina e esse arcabouço é o primeiro grande passo para a estimação computacional de Fluxo Ótico. Vamos falar inicialmente do movimento de corpos rígidos, sem deformações, e de como esse movimento atua sobre a retina.

Podemos, sem perda de generalidade, nos referir à retina como um plano no espaço OXYZ, distante de uma distância z_0 da origem O. A figura 16 mostra como podemos representar geometricamente a configuração do olho. Considere o eixo OZ como a linha de visão e o plano $Z = f$ como o plano focal.

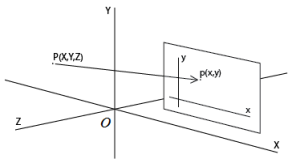

Figura 16: Representação cartesiana do plano da retina.
Fonte: Adaptação de Longuet-Higgins e Prazdny (1980).

Se considerarmos o movimento rígido, ou seja, sem deformações, podemos resumí-lo através de duas componentes: translação e rotação. Por definição, utilizaremos a seguinte notação quando for referido movimento relativo a algum eixo:

- Translação no eixo X, representada por V_X
- Translação no eixo Y, representada por V_Y
- Translação no eixo Z, representada por V_Z
- Rotação sobre o eixo X, representada por W_X
- Rotação sobre o eixo Y, representada por W_Y
- Rotação sobre o eixo Z, representada por W_Z

Portanto, isso implica dizer que $\mathbf{V} = (V_X, V_Y, V_Z)$ é a *velocidade de translação* e $\mathbf{W} = (W_X, W_Y, W_Z)$ é a *velocidade de rotação*.

Suponha um ponto P no espaço, com coordenadas (X_P, Y_P, Z_P) e que a origem O se movimenta no espaço com as velocidades $\mathbf{V}$ e $\mathbf{W}$ descritas acima. Podemos escrever a velocidade de cada componente (X_P, Y_P, Z_P) ponto P como uma composição da velocidade de translação com as respectivas componentes de rotação (LONGUET-HIGGINS; PRAZDNY, 1980).

$$\begin{aligned} \dot{X}_P &= -V_X - W_Y Z_P + W_Z Y_P \\ \dot{Y}_P &= -V_Y - W_Z X_P + W_X Z_P \\ \dot{Z}_P &= -V_Z - W_X Y_P + W_Y X_P \end{aligned} \tag{64}$$

Dessa forma, se projetarmos esses valores e a localização do ponto P sobre o Plano Focal, obtemos a velocidade e a localização na retina, segundo o nosso modelo. A projeção da localização $P = (X_P, Y_P, Z_P)$ sobre (x, y) é

$$(x_P, y_P) = (\frac{X_P}{Z_P}, \frac{Y_P}{Z_P}) \tag{66}$$

e a velocidade pode ser projetada derivando-se a posição (x_P, y_P) tal que

$$\begin{aligned} \dot{x}_P &= \frac{\dot{X}_P}{Z_P} - \frac{X_P \dot{X}_P}{Z_P^2} \\ \dot{y}_P &= \frac{\dot{Y}_P}{Z_P} - \frac{Y_P \dot{Z}_P}{Z_P^2} \end{aligned} \tag{67}$$

ou, utilizando e equação 64, segue-se que

$$\begin{aligned} \dot{x}_P &= (-\frac{V_X}{Z_P} - W_Y + W_Z y_P) - x_P(-\frac{V_Z}{Z_P} - W_X y_P + W_Y x_P) \\ \dot{y}_P &= (-\frac{V_Y}{Z_P} - W_Z x_P + W_X) - y_P(-\frac{V_Z}{Z_P} - W_X y_P + W_Y x_P) \end{aligned} \tag{69}$$

A equação 69 apresenta as velocidades no Plano Focal para as coordenadas x e y, como uma composição das componentes de translação e rotação, que podem ser descritos da seguinte forma:

$$\begin{aligned} \dot{x}_p^{Tra} &= (-V_X + xV_Z)/Z \\ \dot{y}_p^{Tra} &= (-V_Y + yV_Z)/Z \\ \dot{x}_p^{Rot} &= -W_Y + W_Z y + W_X xy - W_Y x^2 \\ \dot{y}_p^{Rot} &= -W_Z x + W_X + W_X y^2 - W_Y xy \end{aligned} \tag{71}$$

onde $\mathbf{v} = (\dot{x}_p, \dot{y}_p) = \mathbf{v}^{Tra} + \mathbf{v}^{Rot} = (\dot{x}_p^{Tra}, \dot{y}_p^{Tra}) + (\dot{x}_p^{Rot}, \dot{y}_p^{Rot})$

Observando as equações em 71, podemos considerar as entidades $x_{FOE} = V_X/V_Z$ e $y_{FOE} = V_Y/V_Z$, de forma que as componentes translacionais podem ser reescritas como

$$\begin{aligned} \dot{x}_p^{Tra} &= (x - x_{FOE})V_Z/Z \\ \dot{y}_p^{Tra} &= (y - y_{FOE})V_Z/Z \end{aligned} \tag{73}$$

que leva a uma relação linear entre $\dot{x}_p^{Tra}$ e $\dot{y}_p^{Tra}$ da forma

$$\frac{\dot{x}_p^{Tra}}{\dot{y}_p^{Tra}} = \frac{x - x_{FOE}}{y - y_{FOE}} \tag{75}$$

Essa relação mostra que a componente translacional de todos os pontos de uma imagem cruzam-se sobre um ponto $\mathbf{x}_{FOE} = (x_{FOE}, y_{FOE})$, o qual Gibson denominou *Foco de Expansão Ótica*, ou somente *Foco de Expansão*, FOE.

A definição do Foco de Expansão é a base para a segmentação do movimento baseado em Fluxo Ótico neste trabalho. O Fluxo Ótico gerado por *egomotion* gera vetores de velocidade cujas componentes translacionais encontram-se no Foco de Expansão. Dessa forma, qualquer vetor de velocidade que fuja à essa regra pode ser encaixado como um objeto independente no meio, conforme será abordado no capítulo 5.

4 TÉCNICAS PARA CÁLCULO DE ANÁLISE DE COMPONENTES INDEPENDENTES

A Análise de Componentes Independentes é um paradigma para a realização de separação cega de fontes, no caso específico em que se supõe que as componentes - ou as fontes - são estatisticamente independentes e linearmente combinadas.

Seja uma variável aleatória X, com distribuição de probabilidade $p_1(X)$, e uma variável aleatória Y com distribuição de probabilidade $p_2(Y)$. Diz-se que essas variáveis aleatórias são independentes se, e somente se, as ocorrências de uma não influenciam nas ocorrências da outra. Isso que dizer que, se existe uma distribuição de probabilidade conjunta $p_3(X,Y)$, vale a relação de independência

$$p_3(X,Y) = p_1(X)p_2(Y)$$

e por isso dá-se a esse paradigma o nome de Análise de Componentes Independentes. Ao longo desde capítulo, será referenciada como ICA, da nomeclatura em inglês, conforme presente na literatura.

As técnicas de ICA surgiram dentro do contexto da medicina, na análise de dados da contração muscular estimulada (HERAULT; JUTTEN, 1991). O problema surgiu através da hipótese de que o sistema nervoso humano seria capaz de alterar a posição angular e a velocidade de uma articulação dado um conjunto de contrações musculares. Por exemplo, quando se dobra um braço, um sinal nervoso é capaz de estimular a musculatura para que as articulações do braço adquiram velocidade e posição angular necessárias para realizar o movimento. Se, de alguma forma, o sistema nervoso humano era capaz de tal realização, esse comportamento poderia ser reproduzido em um ambiente matemático ou, pelo menos, representado por um modelo.

A primeira solução apresentada por Herault et al. (1985) utilizava dois sensores de medição de contração muscular para tentar obter os dados de posição angular e velocidade. Então, o sistema compunha-se de dois sinais observados x_1 e x_2 e dois sinais desejados s_1 e s_2

e, de alguma forma, a composição dos sinais desejados resulta nos sinais observados. Dessa forma, ele pode ser representado por meio de um sistema linear:

$$\begin{aligned} x_1 &= a_{11}s_1 + a_{12}s_2 \\ x_2 &= a_{21}s_1 + a_{22}s_2 \end{aligned} \tag{76}$$

ou, de forma matricial

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{78}$$

onde a_{ij} são os coeficientes da matriz de mixagem $\mathbf{A}$ ou os pesos das componentes s_1 e s_2.

Observe que a linearização do sistema é apenas um modelo, dado que tanto os sinais s_1 e s_2 quanto os coeficientes a_{ij} são desconhecidos. O que Hérault e Jutten queriam era obter os sinais s_1 e s_2 a partir de x_1 e x_2. Supõe-se, então, que o sistema inverso daria uma solução para o problema:

$$\begin{aligned} s_1 &= w_{11}x_1 + w_{12}x_2 \\ s_2 &= w_{21}x_1 + w_{22}x_2 \end{aligned} \tag{79}$$

$$\mathbf{s} = \mathbf{W}\mathbf{x} \tag{81}$$

onde $\mathbf{W} = \mathbf{A}^{-1}$.

Portanto, é necessário que a matriz de mixagem $\mathbf{A}$ seja inversível a fim de que o problema tenha solução. Segundo Hyvärinen et al. (2001), se assumirmos cada coeficiente a_{ij} suficientemente diferentes um do outro, podemos assumir que a matriz de mixagem é inversível.

A primeira abordagem que os autores sugeriram era a de que as componentes desejadas s eram estatisticamente independentes. Dessa forma, eles propuseram uma rede neural retroalimentada simples (figura 17) que resume o algoritmo que posteriormente foi conhecido como *Algoritmo de Hérault-Jutten*, na qual $\hat{s}_i$ é o valor estimado de s_i.

Da figura 17, tiramos que as estimativas $\hat{s}_1$ e $\hat{s}_2$ são:

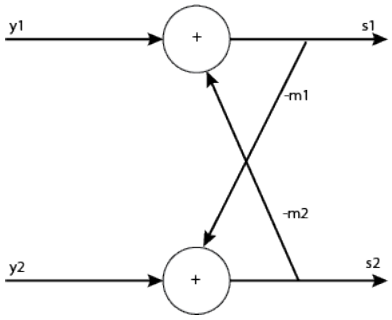

Figura 17: Diagrama do algoritmo de Herault-Jutten.
Fonte: Adaptação de Hyvärinen et al. (2001).

$$\begin{aligned} \hat{s}_1 &= x_1 - w_{12}\hat{s}_2 \\ \hat{s}_2 &= x_2 - w_{21}\hat{s}_1 \end{aligned} \tag{82}$$

$$\hat{\mathbf{s}} = \mathbf{x} - \mathbf{W}\hat{\mathbf{s}} \tag{84}$$

ou, melhor

$$\hat{\mathbf{s}} = (\mathbf{I} + \mathbf{W})^{-1}\mathbf{x} \tag{85}$$

onde $\mathbf{I}$ é a matriz identidade.

A rede, então, deveria adaptar-se de modo que as estimativas $\hat{s}$ fossem estatisticamente independentes entre si. Para tanto, os autores usaram o critério de descorrelação não-linear como forma de medir a independência.

Existe um porém nessa formulação, até o momento. Podemos dizer que variáveis independentes estatisticamente também são descorrelacionadas, mas não o contrário, ou seja, inferir a indepenência estatística baseado em descorrelação. Ao assumir que as variáveis são descorrelacionadas não-linearmente, ou seja

$$E\{f(\hat{s}_1)g(\hat{s}_2)\} = 0 \tag{86}$$

é necessário que algumas restrições sejam impostas para que se possa afirmar a independência de $\hat{s}_1$ e $\hat{s}_2$. Uma vez que f e g são funções quaisquer, é necessario que a equação 86 satisfaça

$$E\{f(\hat{s}_1)g(\hat{s}_2)\} = E\{f(\hat{s}_1)\}E\{g(\hat{s}_2)\} \tag{87}$$

para todas as funções f e g que não sejam nulas em algum intervalo finito. Se assumirmos essas funções como deriváveis em todas as ordens na vizinhança próxima à origem, podemos expandí-las por Taylor, de forma que

$$\begin{aligned} f(\hat{s}_1) &= \sum_{i=0}^{\infty} f^{(i)} \hat{s}_1^i \\ g(\hat{s}_2) &= \sum_{i=0}^{\infty} g^{(i)} \hat{s}_2^i \end{aligned} \tag{88}$$

Assim, o valor esperado do produto dessas séries resume-se a

$$E[f(\hat{s}_1)g(\hat{s}_2)] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f^{(i)} g^{(j)} E[\hat{s}_1^i \hat{s}_2^j] = 0 \tag{90}$$

Essa equação é satisfeita se $E[\hat{s}_1^i \hat{s}_2^j] = 0$ para todo i, j ou se $\hat{s}_1$ e $\hat{s}_2$ são independentes e pelo menos uma das variávels tem média zero. Isso só é possivel se a função com média zero é uma função ímpar, ou seja, a expansão de Taylor só contém as potências ímpares. A presença de potências pares implicaria que os momentos pares, como a variância, devem ser zero, o que só vale se a variável for constante.

Com essas restrições, os autores sugeriram uma regra de aprendizado para a equação 85:

$$\begin{aligned} \Delta m_{12} &= \mu f(\hat{s}_1) g(\hat{s}_2) \\ \Delta m_{21} &= \mu f(\hat{s}_2) g(\hat{s}_1) \end{aligned} \tag{91}$$

onde μ é uma taxa de aprendizado, $f(s) = s^3$ e $g(s) = arctan(s)$, ambas funções ímpares.

Computacionalmente, o algoritmo de Hérault-Jutten não dá garantias de convergência, é pouco robusto a ruído, é pesado e depende fortemente da estimação inicial de $\mathbf{W}$. Entretanto, é de interesse histórico, uma vez que é a partir desse algoritmo que surgiu o modelo ICA.

4.1 A GENERALIZAÇÃO DO CONCEITO DE ICA E AS SUAS RESTRIÇÕES

A equação 78 usada por Hérault e Jutten foi definida para um caso em que haviam duas variáveis a serem determinadas e dois sinais observados. Seja o caso mais genérico, em que n variáveis aleatórias $s_{1,2,3...,n}$ combinam-se linearmente para modelar um sinal observado x_i

$$x_i = \sum_{j=1}^{n} a_{ij} s_j \tag{93}$$

onde, por definição, cada variável s_j é estatisticamente independente de s_k, com $j \neq k$ e os valores de $\{a_{ij}\}$ são coeficientes reais. Nesse caso, tanto os valores de s_j quanto dos coeficientes a_{ij} são desconhecidos e x_i é uma variável observável e mensurável. Esse é o modelo básico de ICA.

Podemos considerar ICA como uma generalização dos métodos de *Separação Cega de Fontes*. Por causa disso, a sua teoria é limitada sob algumas restrições.

Primeiramente, é necessário que as componentes s_j sejam independentes entre si e não possuam distribuições gaussianas. Se a independência não puder ser assumida, o sinal observado x_i não pode ser assumido como uma composição não linear das componentes s_j, pois assume-se que há redundânca de informação e o próprio modelo da equação 78 torna-se inválido.

A restrição quanto a distribuições gaussianas baseia-se na importância de cumulantes de altas ordens dos sinais observados. Cumulantes são os coeficientes da expansão de Taylor da função

$$f(X) = log(E[e^{tX}])$$

para uma variável aleatória X com valor esperado $E[X]$. Assim, o cumulante de primeira ordem, representado por κ_1 é a média e o de segunda ordem, representado por κ_2 é a variância.

As técnicas clássicas de ICA, principalmente as utilizadas neste trabalho, valem-se dos cumulantes de terceira e quarta ordens, que não existem em distribuições gaussianas. Além disso, por questões computacionais, consideramos que o branqueamento, ou seja, a subtração de cada valor indivudual pela média, dos dados é o primeiro passo para qualquer algoritmo de ICA. Dessa forma, não só a complexidade computacional é reduzida - tornando os algoritmos mais rápidos - como algumas condições podem ser assumidas. Por exemplo, sejam os sinais $x_1, x_2, ..., x_n$ representados pelo vetor $\mathbf{x}$. Realizar normalização é transformar o vetor $\mathbf{x}$ em um

vetor $\mathbf{z}$ tal que $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$. Isso pode ser feito por meio de uma matriz $\mathbf{V}$

$$\mathbf{z} = \mathbf{V}\mathbf{x} \tag{94}$$

Essa operação é sempre possível e a literatura a apresenta por vezes com o nome de *sphering* ou transformação esférica. A literatura apresenta várias formas de normalização, como em Bishop (2006) e Trucco e Verri (1998). Pode ser conseguido via PCA, por exemplo, ou através da decomposição EVD da matriz de covariância. Esta forma é convenientemente mais simples do que aquela e baseia-se em reescrever a matriz de covariância $E\{\mathbf{x}\mathbf{x}^T\}$ no produto da matriz de seus autovetores $\mathbf{E}$ e da matriz diagonal formada pelos autovalores $\mathbf{D} = diag(av_1, ..., av_n)$, representado pela equação (95).

$$E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{E}\mathbf{D}\mathbf{E}^T \tag{95}$$

Uma vez realizada essa decomposição, a matriz de *whitening*, $\mathbf{V}$ pode ser estimada como

$$\mathbf{V} = \mathbf{E}\mathbf{D}^{-\frac{1}{2}}\mathbf{E}^T \tag{96}$$

onde, convenientemente, a matriz de autovetores $\mathbf{E}$ é ortogonal e o operador

$$\mathbf{D}^{-\frac{1}{2}} = diag(av_1^{-\frac{1}{2}}, ..., av_n^{-\frac{1}{2}})$$

Sendo esse o caso, com $\mathbf{x} = \mathbf{A}\mathbf{s}$, então

$$\mathbf{z} = \mathbf{V}\mathbf{x} = \mathbf{V}\mathbf{A}\mathbf{s} = \tilde{\mathbf{A}}\mathbf{s} \tag{97}$$

onde $\tilde{\mathbf{A}}$ é ortogonal.

Seja, agora, a distribuição de probabilidade conjunta dos componentes independentes s_j

$$p(s_1, ..., s_n) = \frac{1}{2\pi} exp\left(-\frac{\|\mathbf{s}\|^2}{2}\right) \tag{98}$$

A partir da eq. 98, podemos estimar a probabilidade conjunta dos sinais observados x_i

$$p(x_1,...,x_n) = \frac{1}{2\pi} exp\left(-\frac{\|\mathbf{A^T x}\|^2}{2}\right) |det\mathbf{A}^T| \tag{99}$$

com $\mathbf{A}^{-1} = \mathbf{A}^T$, dada a ortogonalidade de $\mathbf{A}$.Essa condição nos leva a $\|\mathbf{A}^T\mathbf{x}\|^2 = \|\mathbf{x}\|^2$ e $|det\mathbf{A}| = 1$ e, portanto,

$$p(x_1,...,x_n) = \frac{1}{2\pi} exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right) \tag{100}$$

Esse resultado mostra que a utilização de uma matriz de mixagem $\mathbf{A}$ ortogonal sobre $\mathbf{s}$ não altera a distribuição de probabilidade. A matriz $\mathbf{A}$ não pode ser inferida a partir de $\mathbf{x}$, o que também invalida o modelo da eq. 78.

Por fim, restringimos a utilização de ICA para os casos em que o número de sinais observados é o mesmo número de componentes indepentendes. Assim, convenientemente, a matriz de mixagem $\mathbf{A}$ é quadrada. Essa restrição leva a outra, de que $\mathbf{A}$ é inversível. Caso contrário, haveria alguma redundância dentro da matriz que poderia ser omitida e ela deixaria de ser quadrada.

Além dessas restrições, o modelo ICA apresentas algumas limitações. Como ambos $\mathbf{s}$ e $\mathbf{A}$ são desconhecidos, o valor estimado de $\mathbf{s}$ pode vir multiplicado por um escalar qualquer. Esse escalar é compensado, pois a coluna correspondente da matriz de mixagem acaba sendo dividida pelo mesmo escalar, no momento da estimação. Assim, não se pode garantir nem a energia do componente independente, nem o seu sinal. O ICA não é aplicável a sistemas nos quais essas informações (energia e sinal) são relevantes.

Da mesma forma, uma alteração na ordem da soma dos componentes independentes na equação 93 não altera o resultado final. Portanto, não se pode garantir a ordem das ICs.

Casos clássicos, como o *Cocktail Party Problem* não podem ser totalmente resolvidos apenas com técnicas de ICA. Algumas suposições e considerações devem ser feitas para contornar essas limitações. Normalmente, considera-se que cada componente independente possui variância unitária.

Apesar dessas restrições e limitações, as técnicas de ICA apresentam uma gama bem variada de aplicações, tanto no meio acadêmico quanto no meio corporativo. Alguns exemplos de aplicações de ICA, segundo Oja e Hyvärinen (OJA; HYVARINEN, 1997):

- Retirada de ruído de imagens
- Estimação de bases a partir de imagens

- Extração de características a partir do subspaço de cores
- Identificação de artefatos em eletroencefalogramas.
- Separação cega de fontes de canais CDMA convoluídos.
- Aplicações financeiras de modelagem de dados.

4.1.1 OS PRINCÍPIOS DE ESTIMAÇÃO ICA

O conceito de ICA utiliza como princípio a independência estatística, que é uma informação mais "forte" entre variáveis aleatórias do que a descorrelação. Essa afirmação se sustenta no fato de que a independência estatística entre variáveis aleatórias implica que as mesmas também são descorrelacionadas. Mais ainda, implica que transformações não lineares sobre cada variável aleatória mantenham a descorrelação. Esse fato é o primeiro princípio de estimação de ICA, segundo o qual se uma matriz de mixagem **A** for corretamente encontrada, de forma que as variáveis estimadas $\hat{s}_j$ e $\hat{s}_k$, com $j \neq k$ sejam descorrelacionadas *e* as transformações não-lineares por funções ímpares $f(\hat{s}_j)$ e $g(\hat{s}_k)$ também o sejam, as componentes independentes serão encontradas.

Como o modelo ICA baseia-se na soma de variáveis aleatórias não gaussianas, podemos inferir, pelo Teorema do Limite Central (CAM, 1986) que o valor observado **x** é mais gaussiano que qualquer componente independente individualmente. Portanto, o segundo princípio de estimação ICA diz que se for possível encontrar um valor estimado $\hat{s}_j$ que seja o menos gaussiano possível, então encontrou-se um componente independente.

A proxima seção deste capítulo vai explorar ambos os princípios de estimação de ICA. Uma ênfase maior será dada à técnica de estimação de ICA por maximização das características não gaussianas, pois é a solução que implementa o algoritmo FastICA, utilizado neste trabalho.

4.2 ICA POR MAXIMIZAÇÃO DAS CARACTERÍSTICAS NÃO GAUSSIANAS

O segundo princípio de estimação de ICA diz que quanto mais distante a distribuição de probabilidade de um sinal estimado for de uma Gaussiana, mas proximo ele é de um componente independente. Entretanto, precisamos mensurar o quanto uma distribuição é semelhante, ou diferente, de uma distribuição Gaussiana, a fim de realizar a estimação por ICA. Seja o modelo básico apresentado na equação 78. Cada componente independente estimado individualmente representa uma variável aleatória, digamos $\hat{s}_j$ tal que

$$\hat{s}_j = \mathbf{b}^T \mathbf{x} \tag{101}$$

Se **b** corresponder a uma coluna da matriz $\mathbf{a}^{-1}$, inversa da matriz de mixagem, então $\hat{s}_j$ corresponde a um componente independente.Ou seja, o problema da estimação ICA é achar um vetor **b** tal que a multiplicação $\mathbf{b}^T\mathbf{x}$ seja o menos Gaussiano possível. O primeiro ponto a ser considerado é: como medir o quanto Gaussiana uma distribuição é?

4.2.1 A CURTOSE COMO MEDIDA DE CARACTERÍSTICA NÃO GAUSSIANA

Alguns autores, como Cardoso (1989) e Lacoume (1995), propõe a utilização de cumulantes de alta-ordem. A medida mais aceita na literatura é o cumulante de quarta ordem, a *curtose*. A *curtose* de uma variável x pode ser definida como:

$$kurt(x) = E\{x^4\} - 3(E\{x^2\})^2 \tag{102}$$

Esse valor é sempre zero para uma distribuição Gaussiana. Para as demais distribuições, ela pode assumir valores positivos, que classificam a distribuição como supergaussiana conforme a figura 18, ou valores negativos, que classificam a distribuição como subgaussiana (Figura 19).

Na pratica, o valor absoluto da Curtose é utilizado como medida da característica não-Gaussiana.

Algumas características da Curtose devem ser levadas em consideração, dadas as considerações iniciais feitas sobre as técnicas de ICA. Quando consideramos o processo de *whitening*, estamos diretamente afirmando que a variável aleatória com a qual vamos trabalhar tem média zero e variância unitária. Assim, a equação da curtose em 103 pode ser simplificada como

$$kurt(x) = E\{x^4\} - 3 \tag{103}$$

remetendo ao cálculo puro do quarto momento da variável x.

A Curtose também apresenta características lineares que a tornam convenientemente adequadas para a utilização na estimação ICA. Sejam x_1 e x_2 duas variáveis aleatórias independentes e α e β constantes reais, então a função curtose satisfaz a seguinte condição:

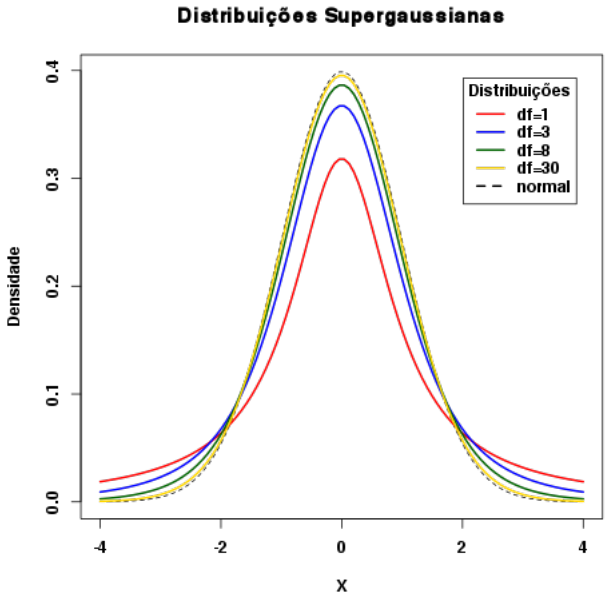

Figura 18: Distribuição supergaussiana
Exemplo de distribuição supergaussiana.
Fonte: Autoria própria.

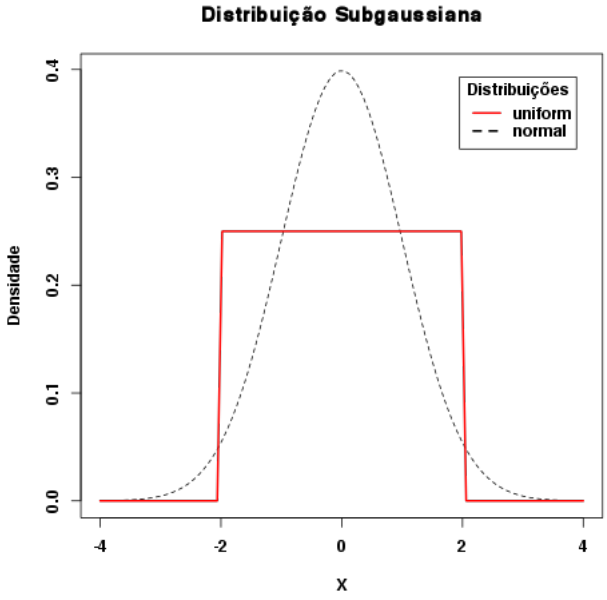

Figura 19: Distribuição subgaussiana
Exemplo de distribuição subgaussiana.
Fonte: Autoria própria.

$$kurt(\alpha x_1 + \beta x_2) = \alpha^4 kurt(x_1) + \beta^4 kurt(x_2) \tag{104}$$

Seja, agora, um vetor de sinais observados $\mathbf{x}$ e uma variável aleatória y tal que

$$y = \mathbf{b}^T \mathbf{x} \tag{105}$$

Se o valor de $\mathbf{b}^T$ for tal que faça o valor absoluto de $kurt(y)$ ser máximo, então y corresponde a uma componente independente (HYVÄRINEN et al., 2001). Considere agora a seguinte transformação:

$$y = \mathbf{b}^T \mathbf{x} = \mathbf{b}^T \mathbf{A}\mathbf{s} = \mathbf{q}^T \mathbf{s} = \sum q_j s_j \tag{106}$$

onde $\mathbf{q} = \mathbf{A}^T \mathbf{b}$.

Se as propriedades lineares da curtose forem aplicadas nesse caso, então

$$kurt(y) = \sum q_j^4 kurt(s_j) \tag{107}$$

Uma otimização proposta é considerar y como tendo variância unitária. Dessa forma, restringimos o valor de q ao círculo n-dimensional unitário, de forma que

$$E\{y^2\} = \sum q_j^2 = 1 \tag{108}$$

Dado que y é a estimação de um dos componentes independentes, é de se esperar que o valor absoluto da curtose seja um máximo local. Prova-se (HYVÄRINEN et al., 2001) que o valor máximo para $kurt(y)$ em uma esfera n-dimensional é quando exatamente um dos valores de q_j é 1 e os demais são 0, ou seja, quando y corresponde à componente independente s_j.

Essa consideração, juntamente com a normalização inicial, é interessante no sentido de que, para um sinal "branco" $\mathbf{z} = \mathbf{A}\mathbf{x}$, estimar ICA é encontrar o vetor $\mathbf{w}$ (que é uma coluna da matriz $\mathbf{A}^{-1}$) tal que se encontre o máximo do valor absoluto da curtose de $\mathbf{w}^T \mathbf{z}$. Vale, portanto, a relação

$$\|\mathbf{q}\|^2 = \|(\mathbf{V}\mathbf{A})^T \mathbf{w}\|^2 = (\mathbf{w}^T \mathbf{V}\mathbf{A})(\mathbf{A}^T \mathbf{V}^T \mathbf{w}) = \|\mathbf{w}\|^2 \tag{109}$$

A grosso modo, a condição apresentada na equação 109 diz que $\mathbf{w}$ também encontra-se

na esfera n-dimensional de raio unitário. Essa informação é a base do algoritmo de estimação ICA por curtose que se segue.

Para todos os algoritmos ICA, o primeiro passo é sempre o processo de branqueamento. Esse processo vai garantir a condição de média nula e variância unitária. A maximização do valor absoluto da curtose pode ser obtida através de um processo iterativo, muito parecido com o algoritmo de Newton-Rhapson. O gradiente do valor absoluto de $kurt(\mathbf{w}^T\mathbf{z})$ pode ser definido como

$$\frac{\partial|\text{kurt}(\mathbf{w}^T\mathbf{z})|}{\partial\mathbf{w}} = 4\text{sign}(\text{kurt}(\mathbf{w}^T\mathbf{z}))[E\{\mathbf{z}(\mathbf{w}^T\mathbf{z})^3\} - 3\mathbf{w}\|\mathbf{w}\|^2] \tag{110}$$

Essa definição nos leva ao primeiro algoritmo de estimação ICA por maximização da curtose por meio do gradiente

Algoritmo 3 Algoritmo de Maximização da curtose por Gradiente

$\mathbf{w} \leftarrow \mathbf{w}_0$
inicializar um fator multiplicativo μ
while $\Delta\mathbf{w} > \varepsilon$ **do**
  $\Delta\mathbf{w} \leftarrow \mu\text{sign}(\text{kurt}(\mathbf{w}^T\mathbf{z}))E\{\mathbf{z}(\mathbf{w}^T\mathbf{z})^3\}$
  $\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$
  $\mathbf{w} \leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}$
end while

Convém observar que se o algoritmo for estável, o gradiente equivale ao vetor $\mathbf{w}$, com mesma direção, multiplicado por um escalar qualquer. Dessa forma, somar $\mathbf{w}$ ao gradiente não altera a sua direção, somente a sua magnitude. Esse fato pode ser utilizado para se definir um novo algoritmo, dito de ponto-fixo, que constitui as bases do algoritmo FastICA. Esse algoritmo converge mais rápido que a abordagem puramente por gradiente (OJA; HYVARINEN, 1997) e não utiliza taxa de aprendizado. Para aplicações mais dinâmicas, é menos custoso computacionalmente, o que o qualifica como mais apropriado. Essas considerações se convertem no algoritmo (4).

Algoritmo 4 Algoritmo FastICA baseado em curtose

$\mathbf{w} \leftarrow \mathbf{w}_0$
while $\Delta\mathbf{w} > \varepsilon$ **do**
  $\mathbf{w} \leftarrow E\{\mathbf{z}(\mathbf{w}^T\mathbf{z})^3\} - 3\mathbf{w}$
  $\mathbf{w} \leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}$
end while

4.2.2 UTILIZAÇÃO DA NEGENTROPIA COMO MEDIDA DE CARACTERÍSTICA NÃO GAUSSIANA

A curtose, como forma de medir a característica não Gaussiana de uma variável aleatória falha quanto à robustez à ruído e à presença de *outliers*. A presença de apenas um resultado fora do esperado ou com ruído causa uma interferência na ordem de $\frac{(valor_{outlier})^4}{N} - 3$, numa sequência de N amostras. Se $valor_{outlier}$ for da ordem de $\sqrt[4]{N}$, então um único valor discrepante dentro de uma sequência pode alterar significativamente o valor da curtose. Sob essa crítica, uma forma mais robusta, porém mais pesada computacionalmente, de se medir a característica não Gaussiana de uma variável aleatória é através da entropia diferencial. A grosso modo, a entropia mede o quanto uma variável é aleatória ou estruturada, em comparação a outras variáveis com a mesma variância. Definimos formalmente a entropia diferencial de uma variável $\mathbf{x}$, com densidade de probabilidades $p_x(\mathbf{x})$ como

$$H(\mathbf{x}) = -\int p_y(\eta) \log p_x(\eta) d\eta \tag{111}$$

Com essa definição, prova-se (PAPOULIS, 1991) que variáveis Gaussianas possuem a maior entropia entre todas as variáveis aleatórias com mesma variância. Essa informação pode ser usada para medir as características não Gaussianas de uma variável aleatória. Seja, então o conceito de negentropia, como a diferença entre a entropia de uma variável Gaussiana e a variável aleatória que se está medindo, de acordo com a equação 112

$$J(\mathbf{x}) = H(\mathbf{x}_{gauss}) - H(\mathbf{x}) \tag{112}$$

A entropia é, por definição, sempre positiva. A negentropia estende esse conceito, de tal forma que $J(\mathbf{x})$ somente será nulo se $\mathbf{x}$ possuir distribuição Gaussiana. Assim, quanto maior o valor da negentropia, menos gaussiana é a variável $\mathbf{x}$.

Na prática, a entropia precisa ser estimada de forma que a sua implementação computacional seja viável. Uma das formas de tornar esse processo viável é através da expansão de Gram-Charlier (HYVÄRINEN et al., 2001) de uma densidade $p_x(\zeta)$ na vizinhança de uma densidade gaussiana. Nesse caso, $p_x(\zeta)$ aproxima-se de uma densidade gaussiana padrão

$$\varphi(\zeta) = exp(-\zeta^2/2)\sqrt{2 * \pi} \tag{113}$$

Dessa forma, podemos reescrever a densidade $p_x(\zeta)$ como

$$p_x(\zeta) \approx \hat{p}_x(\zeta) = \varphi(\zeta)(1 + E\{x^3\}\frac{H_3(\zeta)}{3!} + [E\{x^4\} - 3]\frac{H_4(\zeta)}{4!}) \tag{114}$$

Essa expansão, junto com a aproximação

$$log(1+\varepsilon) \approx \varepsilon - \varepsilon^2/2 \tag{115}$$

para ε pequeno, nos possibilita reescrever a equação da entropia 111 como

$$H(x) \approx -\int \varphi(\zeta) log\,(\varphi(\zeta))\, d\zeta - \frac{(E\{x^3\})^2}{2 \times 3!} - \frac{(E\{x^4\} - 3)^2}{2 \times 4!} \tag{116}$$

Utilizando a equação 116, a equação 112 torna-se

$$J(x) \approx \frac{1}{12} E\{x^3\}^2 + \frac{1}{48} kurt(x)^2 \tag{117}$$

Critica-se o fato de que a equação 117 contenha uma parcela dependente da curtose de x. Logo, sofre das mesmas restrições que o método de maximização da curtose. Para contornar esse problema, uma sugestão é generalizar os cumulantes de alta ordem por funções não quadráticas, eventualmente chamados de momentos não-polinomiais. Assim, os valores de x^3 e x^4 são substituídos por funções $F_1(\cdot)$, par, e $F_2(\cdot)$, ímpar, de modo que a negentropia pode ser aproximada por

$$J(x) \approx k_1(E\{F_1(x)\})^2 + k_2(E\{F_2(x)\} - E\{F_2(\nu)\})^2 \tag{118}$$

onde ν é uma variável Gaussiana padronizada, com média zero e variância unitária.

Costumeiramente, utilizam-se as seguintes funções não polinomiais (HYVÄRINEN et al., 2001)

$$\begin{aligned} F_1(x) &= \frac{1}{a_1} \log \cosh a_1 y \\ F_2(x) &= -exp(-x^2/2) \end{aligned} \tag{119}$$

onde a_1 é um valor real entre 1 e 2.

Na prática, quando x é uma variável aleatória com distribuição simétrica, o primeiro termo da equação 118, $k_1(E\{F_1(x)\})^2$ desaparece. Nesse caso, somente uma função não-

polinomial $F(\cdot)$ é utilizada e essa aproximação (eq. 119) é mais robusta e coerente (HYVÄRINEN et al., 2001). Assim - e considerando a mesma ideia de $\mathbf{w}$ dentro do círculo unitário - obtém-se uma regra de aprendizado para um algoritmo de gradiente 5 usando negentropia:

$$\begin{aligned} \gamma &= E\{F(\mathbf{w}^T\mathbf{z})\} - E\{F(\nu)\} \\ \Delta\mathbf{w} &= \mu\gamma E\{\mathbf{z}G(\mathbf{w}^T\mathbf{z})\} \end{aligned} \tag{121}$$

Algoritmo 5 Algoritmo de gradiente baseado em negentropia

Escolher valores iniciais para $\mathbf{w}$ e γ
while $\Delta\mathbf{w} > \varepsilon$ **do**
  $\Delta\mathbf{w} \leftarrow \gamma\mathbf{z}F(\mathbf{w}^T\mathbf{z})$
  $\mathbf{w} \leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}$
  $\Delta\gamma \leftarrow (F(\mathbf{w}^T\mathbf{z}) - E\{F(\nu)\}) - \gamma$
end while

A mesma consideração de performance sobre o algoritmo de gradiente da curtose vale para o algoritmo 5. O que Hyvärinen et al. (2001) sugere é que a normalização de $\mathbf{w}$ elimina a necessidade de γ. Desse forma, o algoritmo pode ser modificado por um método semelhante ao de Newton, que constitui a implementação do FastICA usando negentropia.

Algoritmo 6 Algoritmo FastICA baseado em negentropia

Escolher valores iniciais para $\mathbf{w}$
while $\Delta\mathbf{w} > \varepsilon$ **do**
  $\mathbf{w} \leftarrow E\{\mathbf{z}F(\mathbf{w}^T\mathbf{z}) - E\{F'(\mathbf{w}^T\mathbf{z})\}\mathbf{w}$
  $\mathbf{w} \leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}$
end while

4.3 ICA COMO FILTRO DE BORDAS

Um dos temas motivadores deste trabalho é a abordagem de Bell e Sejnowski (1997), baseada na afirmação de Barlow (1989) de que a capacidade que o cérebro humano possui de identificar bordas é resultado de um processo de redução de redundância das informações visuais. As informações restantes são, a princípio, independentes umas das outras e, dentro dessas, estão filtros de borda. Os trabalhos de Herault e Jutten (1991) e de Comon (1994) levaram os autores à suposição de que se Barlow estivesse correto, então o cérebro humano realizaria um processo parecido com a estimação de ICA e a aplicação de um de seus algoritmos levaria a filtros de bordas.

O primeiro passo a ser considerado é a modelagem do sistema de acordo com o modelo básico de ICA. O trabalho de Olshausen e Field (1996) demonstra que o sistema perceptual humano está sujeito a vários retalhos de imagens. Cada um desses retalhos é gerado a partir de uma conjunto de imagens base, ou "causas" (BELL; SEJNOWSKI, 1997), cada uma sujeita a uma "função base". Considere que um retalho de imagem é uma região de uma imagem maior. Podemos, então, dizer que é uma matriz de intensidades em que cada elemento corresponde a um *pixel*. Por conveniência, essa matriz pode ser reescrita em um vetor $\mathbf{x}$. As funções-base são da mesma forma imagens e através da mesma analogia podem ser escritas como vetores de intensidades $\mathbf{a}_i$. Assim, podemos agrupar cada uma dessas funções-base como colunas de uma matriz $\mathbf{A}$. Cada uma das funções base possui um peso, s_j.

Essas definições levam ao modelo ICA

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

O que os Bell e Sejnowski esperavam era resolver essa a equação e determinar a matriz $\mathbf{A}$. Cada coluna da matriz corresponderia a um filtro de borda. Para tanto, os autores utilizaram um método de de gradiente que utiliza a entropia conjunta $H[g(\mathbf{W}^{-1}\mathbf{x})]$, onde g é uma função sigmóide. Essa abordagem encontra-se no trabalho de Cardoso e Laheld (1996) e utiliza a função de aprendizado representada na equação 123.

$$\Delta\mathbf{W} = \mu\frac{\partial H(g(\mathbf{s}))}{\partial\mathbf{W}}\mathbf{W}^T\mathbf{W} = (\mathbf{I} + \hat{\mathbf{y}}\mathbf{s}^T)\mathbf{W} \tag{123}$$

onde $\hat{\mathbf{y}}$ é um vetor em que cada elemento i corresponde a uma não-linearidade $\hat{y}_i = \dfrac{\partial}{\partial s_i}\ln\dfrac{\partial g(s_i)}{\partial s_i}$.

Convenientemente, os autores utilizaram uma matriz de retroalimentação $\mathbf{V}$ de forma que

$$\mathbf{s} = (\mathbf{I} + \mathbf{V})^{-1}\mathbf{x} \tag{124}$$

e assim utilizaram uma rede neural semelhante à de Hérault e Jutten da figura 17, com regra de aprendizado

$$\Delta\mathbf{V} = (\mathbf{I} + \mathbf{V})(\mathbf{I} + \hat{\mathbf{y}}\mathbf{s}^T) \tag{125}$$

Com essa rede definida, os autores selecionaram um conjunto de imagens naturais em

escala de cinza, compostas de cenas de paisagens. Várias amostras de tamanho 12×12, ou retalhos, foram aleatoriamente selecionados. Cada uma dessas amostras constitui uma amostra de $\mathbf{x}$, que alimentaram a rede em seu treinamento. A figura 20 mostra os resultados obtidos por Bell e Sejnowski e comparados com outros métodos de retirada de redundância, como análise por PCA e por Análise de componentes de fase zero, ZCA.

É plausivel o pensamento de que o algoritmo apresentado é biologicamente impossível, pois envolve um mecanismo de retroalimentação sem correspondentes do Sistema Visual Humano. Também não conclui afirmativamente, apenas dá indícios, de que o cérebro utiliza ICA. Entretanto, o resultado encontrado é importante, no sentido de entender que realmente existe um processo de redução de redundância nas células do córtex visual. Os indícios até então levam a crer que esse processo é uma forma de ICA, de forma que os filtros de borda presentes no sistema perceptual humano sejam, por si só, independentes.

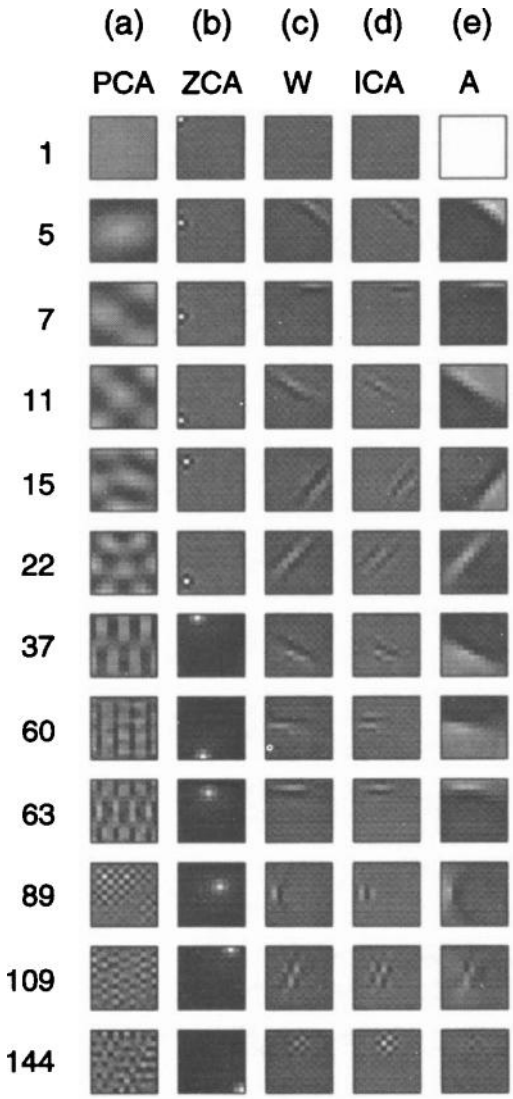

Figura 20: Resultado de Bell e Sejnowski.
Resultado obtido para a aplicação de ICA sobre retalhos de imagens. Cada uma das imagens corresponde a uma coluna da matriz de mistura obtida.
Fonte: Bell e Sejnowski (1997).

5 ICA APLICADA A FLUXO ÓTICO

Para que qualquer técnica de ICA seja aplicada à solução de um problema, é necessário entender quais e quantas são as variáveis aleatórias consideradas, quantas e quais são as componentes independentes e quais são os sinais observáveis plausíveis. Dessa forma, consegue-se modelar os dados segundo a estrutura básica ICA apresentada no capítulo 4. A hipótese inicial deste trabalho é a de que o fluxo ótico gerado pela movimentação de um objeto livre e independente é estatisticamente independente do fluxo ótico gerado pela movimentação própria da câmera (*egomotion*). As bases dessa hipótese partem da lógica de que o Fluxo Ótico é resultado da movimentação relativa entre a câmera e o fundo e objetos com movimentação própria. Assim, podemos considerar que o Fluxo Ótico estimado é uma combinação linear das componentes vetoriais de *egomotion* e dos objetos animados na cena. Essas componentes vetoriais serão as componentes independentes do modelo ICA.

O modelo mais simplificado de ICA supõe uma matriz de mixagem quadrada, de forma que é necessário um número de sinais observáveis igual ao de componentes independentes. A primeira ideia considerada para resolver esse problema foi a utilização de vídeo estéreo, dois vídeos de uma mesma cena, sob duas perspectivas ligeiramente diferentes. Essa primeira tentativa possuía uma restrição muito grande quanto à disponibilidade de vídeos estéreo com componente de *egomotion*. Foram utilizadas sequências de imagens disponíveis na *The Computer Vision Homepage* (HUBER, 2004) e não havia vídeo estéreo disponível no qual ambas as câmeras realizassem o mesmo movimento simultaneamente, apenas vídeos com fundo parado. Nesse caso, esse conhecimento *a priori* de que não há componente de movimento próprio não justificaria a utilização de um algoritmo de ICA, dado que se há vetor de Fluxo Ótico, então ele necessariamente é de um objeto com movimentação independente da câmera.

A modelagem para vídeos estéreo gerou um novo ponto a ser considerado, sendo necessário entender melhor quais eram as variáveis aleatórias que seriam utilizadas como sinais observados. Cogitou-se a utilização dos vetores de movimento do Fluxo Ótico estimado. Entretanto, uma amostragem seria um vetor de movimento em uma posição fixa (x, y) de um quadro ou seria relativo a um ponto de interesse cuja posição muda constantemente ao longo do

vídeo. Essa discussão levaria à escolha do melhor método de estimação de Fluxo Ótico. Para o primeiro caso, uma estimativa densa bastaria. Para o último, seria necessaria uma estimativa esparsa e os pontos de interesse deveriam ser acompanhados ao longo do vídeo. A primeira tentativa realizada foi considerar fluxo ótico esparso, com acompanhamento dos pontos de interesse. Principalmente na hipótese de vídeo estéreo, seria necessário realizar o casamento dos pontos do vídeo da direita com os do vídeo da esquerda e dos pontos de interesse de um quadro com o do quadro subsequente, em ambos.

Essa hipótese era inconveniente no sentido do número de amostras disponíveis. Um vídeo com 10 segundos tem aproximadamente 300 quadros. Dessa forma, seria necessário que o mesmo ponto de interesse pudesse ser identificado em todos os quadros do vídeo para que se tivessem 300 amostras. Convenientemente, uma nova versão da biblioteca OpenCV (OPENCV, 2012), com suporte a SURF (BAY et al., 2008), tinha sido recentemente lançada. Essa primeira hipótese foi avaliada medindo-se a quantidade de quadros consecutivos através dos quais fosse possível identificar um ponto de interesse. O experimento era realizar a estimativa de fluxo ótico com casamento de pontos de interesse através de descritores SURF. Para cada descritor gerado, cada aparição em cada quadro era computada. Dessa maneira e utilizando um vídeo de controle, *car.avi* (STAVENS, 2005), o máximo de quadros consecutivos que um descritor SURF foi identificado foram 55 quadros. Já é um indicativo que a quantidade de amostras que essa abordagem forneceria não seria suficiente para uma técnica ICA, que possui caráter estatístico.

Nesse experimento, a abordagem por fluxo ótico esparso foi inviabilizada. Caberia, então, validar a abordagem via fluxo ótico denso. Nessa abordagem, mantendo-se a mesma grade de vetores de movimento, cada nó corresponderia a uma amostra. Dessa forma, a passagem de um quadro para outro geraria uma quantidade considerável de amostras e não seria necessário fazer casamento entre os pontos de interesse. Até então, não havia vídeos estéreo com a componente de *egomotion*. Era necessário encontrar uma saída alternativa. Os autores Bell e Sejnowski (1997) utilizam uma abordagem interessante para o caso dos filtros de bordas. Eles segmentaram imagens em retalhos de tamanho $12x12$, no qual cada *pixel* corresponde a uma amostragem da um sinal observável para 144 fontes. Essa mesma abordagem poderia ser estendida para os vetores de movimento.

O trabalho com vídeos possui um fator temporal que permite uma liberdade maior na segmentação do campo de Fluxo Ótico. Partindo desse ponto e considerando que se estava trabalhando com duas componentes independentes, foram utilizados retalhos 2×1 sob duas perspectivas.

A primeira segue a lógica da figura 21, na qual a cada par consecutivo de vetores de

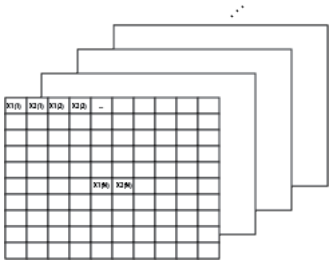

Figura 21: Primeiro modelo experimental.
Cada par de vetores de movimento aproxima o comporamento de duas variáveis aleatórias observáveis.
Fonte: Autoria própria.

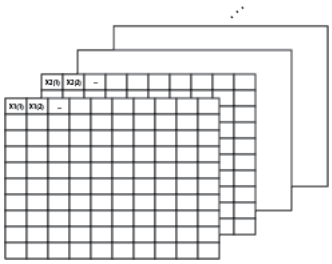

Figura 22: Segundo modelo experimental.
Cada campo de Fluxo Ótico corresponde a um arranjo de variáveis aleatórias observáveis.
Fonte: Autoria própria.

movimento, o primeiro corresponde a uma amostra do vídeo da esquerda e o segundo a uma amostra do vídeo da direita. Esta é uma abordagem espacial.

A segunda segue a lógica da figura 22, na qual para cada par de quadros consecutivos do vídeo, o primeiro quadro corresponde ao quadro da esquerda e o segundo quadro corresponde ao da direita. Assim, cada vetor de movimento corresponde a uma amostragem em ambos os casos. Esta é uma abordagem temporal.

Nesse segundo caso, considera-se ou que o movimento da câmera possui um comportamento contínuo ou que as variações no padrão de movimento são irrelevantes entre dois quadros consecutivos.

Para cada uma dessas perspectivas, cabe observar em que momento do vídeo aplicar o algoritmo de ICA. Uma opção é a cada campo de Fluxo Ótico estimado. O número de amostras utilizado é reduzido, entretanto, não há erros relativos a descontinuidade do movimento da câmera. Uma outra opção é rodar o algoritmo ao final de todas as estimações dos campos de Fluxo Ótico. Assim, o número de amostras é maior, mas supõe-se que o movimento da câmera é constante ao longo de todo o vídeo. Dessa forma, existem quatro modelos experimentais para

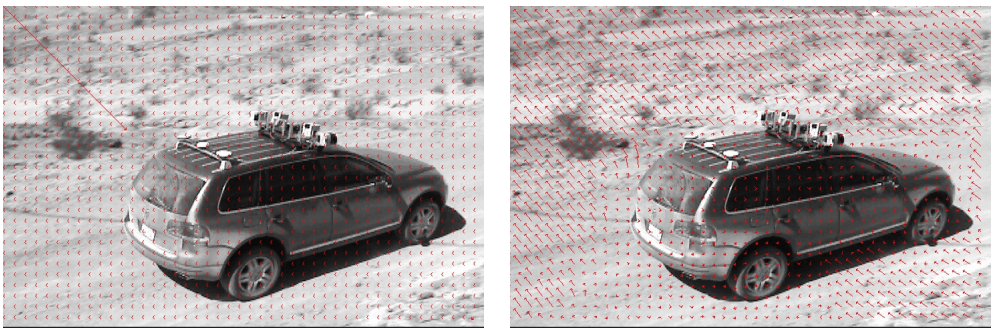

Figura 23: Resultado da aplicação de ICA ao primeiro modelo experimental.
Considerando uma execução do algoritmo de ICA a cada campo de Fluxo Ótico estimado. Utiliza uma quantidade menor de amostras.
Fonte: Autoria própria.

os quais pode-se testar ICA sobre vetores de Fluxo Ótico.

Foi utilizada a biblioteca OpenCV 2.1 para a estimação dos campos densos de Fluxo Ótico via algoritmo de Farnebäck (2002) e uma biblioteca para cálculo chamada IT++ para a separação de componentes independentes via algoritmo FastICA. Convém observar que a implementação do algoritmo FastICA não comporta o cálculo de variáveis aleatórias multi-dimensionais. Portanto, foi necessário desmembrar os vetores de movimento em suas componentes X e Y e para cada uma executar o algoritmo FastICA. O resultado final é a combinação dos resultados encontrados em cada uma dessas execuções.

A ideia é obter duas componentes vetoriais. Uma relativa à movimentação da câmera e outra a de possíveis objetos independentes presentes no vídeo, de forma que cada vetor de movimento possa ser classificado individualmente. As figuras 23, 24, 25 e 26 mostram as componentes de *egomotion* (a) e de objeto independente (b) para a primeira perspectiva (figura 21) calculada a cada campo de Fluxo Ótico calculado, ao final de todos os cálculos, para a segunda perspectiva (figura 22) calculada a cada campo de Fluxo Ótico e ao final de todos os cálculos, respectivamente.

Os resultados com o vídeo de controle indicam uma taxa de erro muito grande e evidenciam a inviabilidade prática. A primeira análise é a de que os modelos aplicados não são os ideais. Uma das características das técnicas de ICA é a necessidade de uma grande quantidade de amostras. Quanto mais amostras, mais precisa é a solução. No caso dos campos de Fluxo Ótico estimados, a quantidade de amostras é restrita, o que, por si só seria uma restrição à utilização de ICA sobre os vetores de movimento.

O que as figuras 23, 24, 25 e 26 apontam é a que as técnicas de ICA buscam pelas

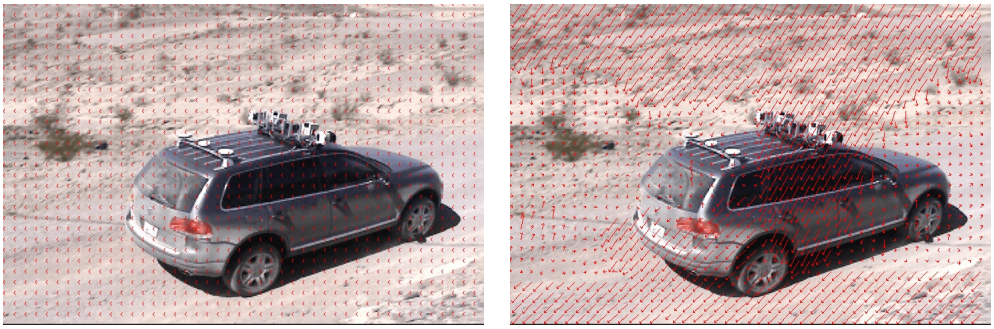

Figura 24: Resultado da aplicação de ICA ao primeiro modelo experimental.
Considerando uma única execução do algoritmo de ICA ao final de todos os Campos de Fluxo Ótico estimados.

Fonte: Autoria própria.

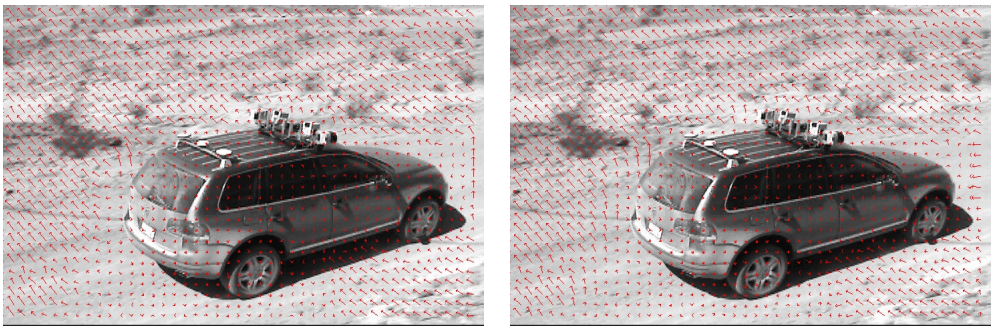

Figura 25: Resultado da aplicação de ICA ao segundo modelo experimental.
Considerando uma execução do algoritmo de ICA a cada campo de Fluxo Ótico estimado.

Fonte: Autoria própria.

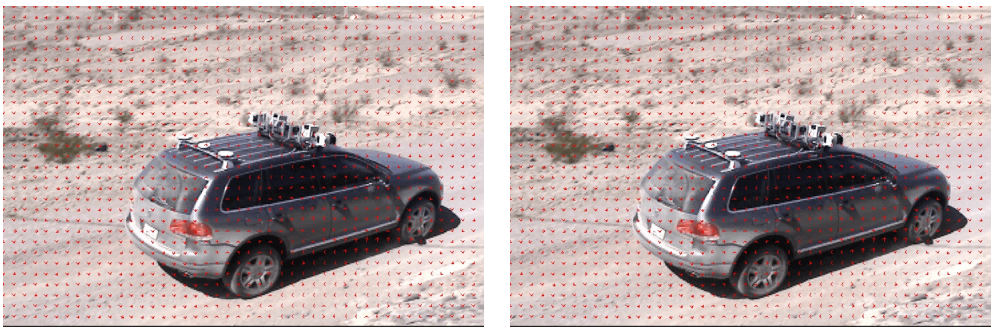

Figura 26: Resultados da aplicação de ICA ao segundo modelo experimental.
Considerando uma única execução do algoritmo de ICA ao final de todos os Campos de Fluxo Ótico estimados.

Fonte: Autoria própria.

amostras de componentes independentes em todas as amostras observáveis. Ou seja, todos os vetores de movimento apresentariam tanto uma componente de fundo quanto uma componente gerada por um objeto animado independente. De acordo com Hyvärinen et al. (2001), isso se dá pois a matriz de mistura assume um valor constante para toda a seqÃ¼ência amostral utilizada. Logo, cada amosta (vetor de movimento) possuiria ambas as componentes independentes. Um vetor de movimento que representa fundo não apresenta uma componente gerada por um objeto animado, portanto, para esse valor, a matriz de mistura não se aplicaria. Seria necessário, então, aplicar os algoritmos de ICA apenas sobre vetores que contivessem ambas as componentes. Essa limitação implica que o conhecimento *a priori* de quais vetores são gerados somente pelo fundo e quais são gerados por um objeto animado é necessário.

Até então, não havia uma base de dados de Fluxo Ótico com gabarito, de forma que esse conhecimento *a priori* fosse possível ou mesmo para que uma análise objetiva fosse realizada. Optou-se por criar um banco de dados com campos de Fluxo Ótico sintéticos, com a presença de objetos animados, em que cada vetor de movimento possui gabarito. Ao mesmo tempo e diante da inviabilidade de utilização de ICA, optou-se por procurar uma forma alternativa de segmentação que seja mais adequada aos modelos de estimação de movimento atuais, baseados na representação vetorial do movimento. Foi escolhida uma abordagem pelo Foco de Expansão (GIBSON, 1979), conforme será visto no capítulo 6.

6 TÉCNICAS ALTERNATIVAS DE SEGMENTAÇÃO

A solução de segmentação de movimento por ICA não se mostrou conveniente nem viável. Era necessário procurar uma solução alternativa. A parametrização do movimento sugere que é possível estimar parâmetros para um campo de Fluxo Ótico e através desses parâmetros classificar cada vetor de movimento. O que torna essa solução difícil é a presença dos objetos com livre movimentação, cujo padrão de movimento é diferente do padrão do fundo. Não há uma solução definitiva para este problema. A literatura apresenta formas de estimação do *egomotion*, como Bruss e Horn (1983), Heeger e Jepson (1992), Tomasi e Shi (1993), Prazdny (1980), Kanatani (1993). A ideia central destes métodos é estimar separadamente as componentes translacionais e rotacionais através da compensação de cada uma dessas componentes. Este capítulo aborda uma forma de classificar os vetores de movimento através da estimação do Foco de Expansão.

6.1 CONSTRUÇÃO DE UM BANCO DE DADOS SINTÉTICO

O campo de Fluxo Ótico pode ser parametrizado e modelado. Considerou-se que o fundo é um objeto rigido, portanto, é coerente pensar que todos os vetores de movimento de *egomotion* seguem o mesmo modelo. Utilizou-se a definição de alguns modelos parametrizados de movimento (STILLER; KONRAD, 1999) para simular o campo de Fluxo Ótico em diversas situações. Cada um dos modelos foi sintetizado sobre imagens de 480×360 *pixels*, com grade de 16 *pixels*. Para este banco de dados, utilizaram-se os modelos de movimento translacional e afim.

Inicialmente, foi simulado apenas movimento translacional, arbitrando-se a localização do foco de expansão e calculando-se os ângulos para cada nó da grade de Fluxo Ótico. Neste caso, considerou-se que focos de expansão com coordenadas X ou Y com valores absolutos maiores que 9999 *pixels* estão no infinito. Para esses casos, criaram-se cenários cujo foco de expansão varia de $(-1000, -1000)$ a $(1000, 1000)$, com intervalos de 75 *pixels* para X e Y de forma que o vetor de movimento tenha amplitude constante de 10 *pixels*. Manter a ampli-

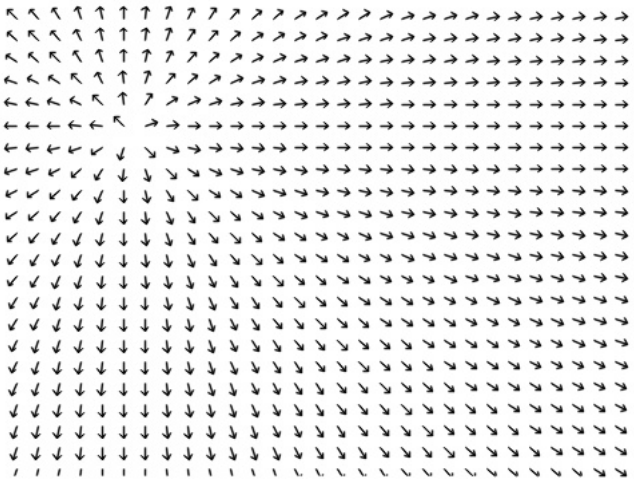

Figura 27: Primeiro exemplo de campo de Fluxo Ótico para modelo Translacional.
Gerado a partir de um Foco de Expansão em (100, 100).
Fonte: Autoria própria.

tude do vetor de movimento constante é uma condição que desconsidera qualquer variação de profundidade na imagem. Entretanto, facilita a criação da base de dados, garante a visualização dos vetores de movimento e possibilita a validação de algoritmos. A escolha do tamanho do intervalo é empírica e serve para restringir o número de arquivos gerados. As figuras 27 e 28 mostram dois exemplos de movimento translacional com focos de expansão em $(100,100)$ e $(1000,1000)$, respectivamente.

Dessa forma, para um foco de expansão em (x_{FOE}, y_{FOE}) cada vetor de movimento $\mathbf{d}(x,y) = (dx,dy)$ tem suas componentes dx e dy calculadas através da equação 126.

$$\begin{aligned} dx &= \frac{10}{\sqrt{1+m^2}} \frac{x - x_{FOE}}{|x - x_{FOE}|} \\ dy &= m \cdot dx \end{aligned} \tag{126}$$

onde $m = \dfrac{y - y_{FOE}}{x - x_{FOE}}$ é o coeficiente linear da reta suporte que liga (x,y) a (x_{FOE}, y_{FOE}).

Depois, simulou-se movimento afim, através da equação paramétrica (128). O que é interessante de ser observado é que simular movimento afim insere uma componente rotacional.

$$\mathbf{d}(x,y) = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_3 \\ b_3 \end{bmatrix} \tag{128}$$

As figuras 29 e 30 mostram dois exemplos de campos de Fluxo Ótico obtidos pela

Figura 28: Segundo exemplo de campo de Fluxo Ótico para modelo Translacional.
Gerado a partir de um Foco de Expansão em (1000, 1000). Para imagens pequenas, essas coordenadas aproximam o infinito.
Fonte: Autoria própria.

Tabela 1: Exemplos de modelo afim parametrizado.
Os valores das variáveis $a1$, $a2$, $b1$ e $b2$ representam a componente de rotação. Já os valores de $a3$ e $b3$ representam as componentes de translação.

Modelo	a1	a2	a3	b1	b2	b3
Modelo 1	-0,1	0,1	-5	0,1	-0,2	10
Modelo 2	-0,1	0,3	-5	0,1	0,4	10

equação (128), para os parâmetros da tabela 1. Para o modelo de movimento afim, manteve-se o valor dos parâmetros $a_3 = 5$ e $b_3 = 10$ constantes, pois são as componentes translacionais. Os valores de a_1 e b_2 foram variados de -1 a 1 em intervalos de 0.1 e os valores de a_2 e b_1 foram variados de -0.1 a 0.1, com intervalos de 0.01.

Convém observar que a ideia principal do modelo afim é observar o comportamento dos algoritmos com componentes translacionais e rotacionais. Os valores rotacionais a_1, a_2, b_1 e b_2 foram propositalmente configurados para serem baixos, pois na época em que esta base de dados estava sendo construída, estava-se trabalhando em uma forma alternativa de segmentação de movimento baseada em foco de expansão, que é sensível a rotação. Com valores baixos de rotação, a influência dessas componentes pode ser mais facilmente vista.

Ainda era necessária a inserção de ao menos um objeto independente em cada um desses campos de Fluxo Ótico modelados. Considerou-se que um objeto independente é uma região no campo vetorial que foge ao comportamento do modelo paramétrico. Assim, não importa o modelo utilizado para o objeto, contanto que ele seja consistente com movimento rígido. Adicionaram-se vetores de movimento que seguem os mesmos modelos paramétricos transla-

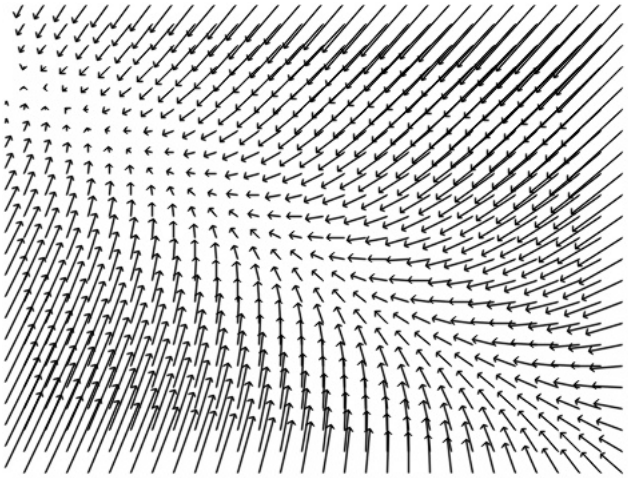

Figura 29: Exemplo de Fluxo Ótico estimado para o primeiro modelo afim.
Utilizando os parâmetros para o Modelo 1 da tabela 1. Estes vetores não estão normalizados.
Fonte: Autoria própria.

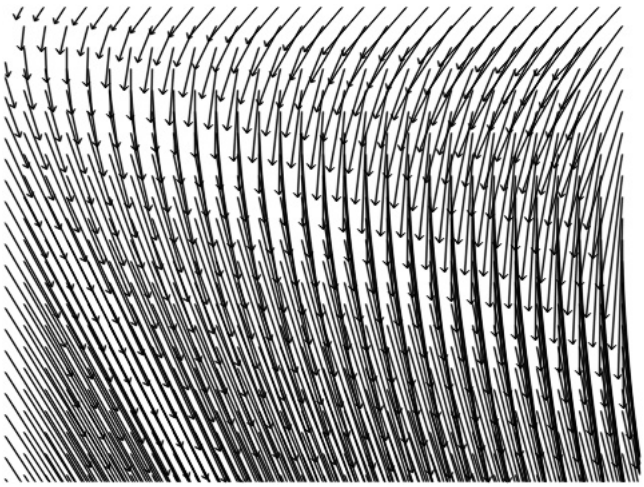

Figura 30: Exemplo de Fluxo Ótico estimado para o segundo modelo afim.
Utilizando os parâmetros para o Modelo 2 da tabela 1. Estes vetores não estão normalizados.
Fonte: Autoria própria.

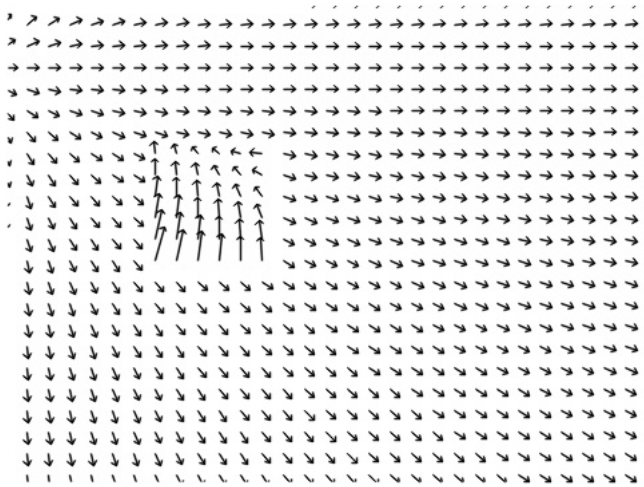

Figura 31: Movimento translacional com objeto.
Modelo gerado a partir de um Foco de Expansão em (-100, 100), considerando um objeto quadrado de lado 150 pixels, com movimento afim.
Fonte: Autoria própria.

cionais ou afins, mas com parâmetros diferentes do modelo de movimentação do fundo, dentro de uma região delimitada por um retângulo ou um círculo com tamanhos variáveis. Assim, ao todo 29282 cenários diferentes para movimento afim e 1458 para movimento translacional foram criados.

As figuras 31 e 32 mostram exemplos dos cenários finais da base de dados, para movimento translacional e afim, respectivamente.

Cada uma dessas imagens foi salva em arquivos, contendo os parâmetros de movimento do fundo, os parâmetros do movimento do objeto, o tamanho da imagem, o espaçamento entre os nós da grade de Fluxo Ótico e as informações dos vetores de movimento. Estas, são constituídas pela posição (x, y) do nó, o seu deslocamento (dx, dy) e um indicativo se este vetor de movimento é correspondente ao fundo (‘b’) ou ao objeto (‘o’).

6.2 SEGMENTAÇÃO POR FOCO DE EXPANSÃO

A primeira coisa a se ter em mente é que o escopo desta abordagem abrange apenas movimentos translacionais. A análise da estimação de focos de rotação é um possível trabalho futuro. Entretanto, uma análise do comportamento dos algoritmos sobre movimento com componente rotacional é realizada.

Esta abordagem só é possível se considerarmos que a maioria de *pixels* ao longo de dois quadros consecutivos de um vídeo é de fundo, ou seja, se há o conhecimento à *priori* de

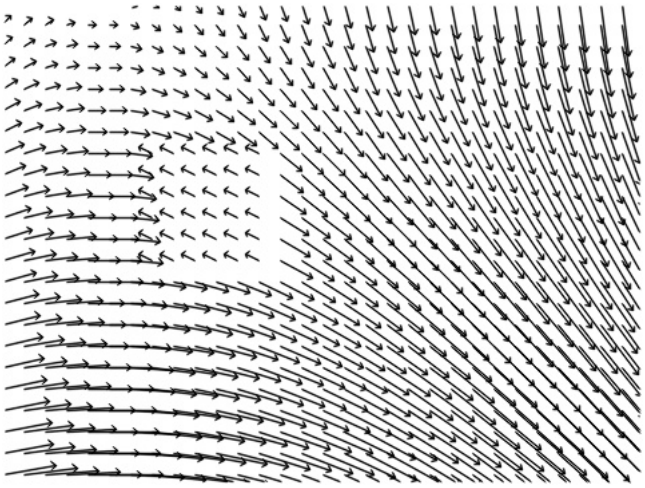

Figura 32: Movimento afim com objeto.
Modelo gerado a partir da equação 128, com um objeto quadrado de lado 150 pixels, com movimento translacional.
Fonte: Autoria própria.

que a maioria dos vetores de movimento no campo de Fluxo Ótico correspondem a vetores de *egomotion*. Dessa forma, pode-se estimar um "comportamento médio"através do qual pode-se estimar um foco de expansão. Cada par de vetores de movimento foi relacionado geometricamente e estimou-se o ponto de encontro das retas de suporte. Esse constitui um foco de expansão relativo. Existem três possíveis focos de expansão relativos:

- Entre dois vetores de movimento de *egomotion*. Nesse caso, ele deve estar espacialmente próximo ao foco de expansão do fundo.
- Entre um vetor de movimento de *egomotion* e um de objeto. Pode estar localizado em qualquer posição espacial. É um *outlier*.
- Entre dois vetores de movimento de objeto. Pode estar localizado em qualquer posição espacial.

Antes de qualquer solução formal, foi realizado um teste preliminar simples, no qual para cada par de vetores de movimento, um foco de expansão relativo foi calculado e armazenado. Dentre todos os focos de expansão estimados, foram contabilizados os iguais, dentro de um pequena faixa de variação e foi assumido como Foco de Expansão de *egomotion* aquele com o maior número de incidências. Com esse Foco de Expansão final, pode-se modelar na imagem o campo de Fluco Ótico e compará-lo ao original. Assim, a classificação se dá através do produto interno entre o vetor de movimento normalizado criado sinteticamente e o vetor

estimado através da abordagem de Foco de Expansão para cada ponto da grade do campo de Fluxo Ótico. Se considerarmos que ambos os vetores estão normalizados, o produto interno entre eles aproxima-se do valor 1, caso os vetores sejam semelhantes e aproximam-se do valor -1 caso sejam inversos. Assim, segue-se o algoritmo de classificação, pela abordagem do produto interno 7.

Algoritmo 7 Classificação dos vetores de movimento pelo Foco de Expansão

$angle_{FOE} \leftarrow atan\left(\frac{y - y_{FOE}}{x - x_{FOE}}\right)$
$dx_{FOE} \leftarrow cos(angle_{FOE})$
$dy_{FOE} \leftarrow sin(angle_{FOE})$
$produtoInterno = dx \times dx_{FOE} + dy \times dy_{FOE}$
if $produtoInterno < \varepsilon$ **then**
  $type \leftarrow' background'$
else
  $type \leftarrow' object'$
end if

onde $angle_{FOE}$ é o angulo formado pela reta suporte que passa pelo Foco de Expansão e o ponto (x,y) que está sendo considerado, (dx_{FOE}, dy_{FOE}) é o vetor de movimento normalizado para o Foco de Expansão estipulado, (dx,dy) é o vetor de movimento real e $\varepsilon = 0,3$ é uma tolerância de erro escolhida empiricamente.

Esta solução obteve uma taxa de acerto para fundo com movimento translacional de 77% e para movimento afim de 57%. A surpresa neste caso é para o movimento translacional. 23% de erro é uma taxa muito alta, considerando que existe o conhecimento prévio de que os vetores são de fundo e o cruzamento das suas retas de suporte deveriam coincidir com o Foco de Expansão através do qual o cenário foi criado. Uma análise mais profunda tanto do processo de criação do banco de dados, quando do processo de classificação revelou que problemas de arredondamento são a primeira causa visível de erros. É preciso considerar que quando se trabalha com coordenadas em *pixels*, trabalha-se com valores inteiros. Foi observado que quanto menor a amplitude do vetor de movimento, mais sensível a esses problemas de arredondamento a solução se torna. Isso, no caso da estimação de Fluxo Ótico em uma situação real, em que há o efeito da profundidade, é crítico.

As tabelas 2 e 3 mostram a taxa de acerto dessa regra de classificação para os modelos translacional e afim. A taxa de acerto é obtida, aplicando-se o algoritmo sobre os arquivos do banco de dados sintético. Cada vetor de movimento é então classificado entre fundo ('b') e objeto ('o') e os resultados são comparados com a base de dados original. Cada acerto e cada erro é computado e considera-se a taxa de acerto como a relação entre o número de acertos e o

número total de comparações.

A tabela 2 apresenta algumas posições do Foco de Expansão (X_{FOE}, Y_{FOE}) e as respectivas taxas de acerto. Por exemplo, para um Foco de Expansão com

$$(X_{FOE}, Y_{FOE}) = (-1000, 1000)$$

a taxa de acerto foi de $0,6348$ ou 63%. Na tabela 3, mostra-se o efeito sobre a taxa de acerto da variação dos parâmetros a_2, b_1 e b_2, mantendo a_1 fixo, conforme a equação 128. Pode-se observar que a taxa de acertos diminui quando aumentam-se os valores das variáveis relativas à componente rotacional.

Para essa análise, os arquivos foram gerados em um MacBook Pro Dual Core, com 4GB de RAM. O processo de análise compunha-se de: criação dos arquivos do banco de dados sintético, estimação do Foco de Expansão para cada arquivo, classificação e avaliação dos resultados. O processo mais custoso computacionalmente nesta abordagem é a estimação do Foco de Expansão. Por ser uma abordagem inocente, a comparação dos vetores de movimentos dois a dois resulta em uma quantidade grande de cálculos realizados. Mesmo com uma base de dados sintética, na qual não há estimação de movimento, o tempo de processamento da estimação do Foco de Expansão ultrapassa os 40 minutos (2451 segundos) para 30740 campos vetoriais simulados.

Já a classificação dos vetores é menos custosa. Se considerarmos o critério de classificação baseado em ângulo, os 30740 campos vetoriais levaram 461 segundos de processamento. Quando consideramos o produto interno dos vetores como critério de classificação, foram 313 segundos. Em uma aplicação em tempo real, soma-se a esse tempo o processamento da estimação de movimento. Utilizando-se o estimador de Farnebäck implementado na biblioteca OpenCV sobre o vídeo de controle, que possui 10 segundos de duração e 304 quadros, somente a estimação de Fluxo Ótico demorou cerca de 61 segundos. Essa análise indica que, mesmo uma abordagem não tão inocente de estimação do Foco de Expansão não seria por si só suficiente para uma situação em tempo real. Convém ressaltar que o objetivo desta pesquisa não era a otimização computacional nem melhorias do desempenho do processamento.

Quando o mesmo experimento foi realizado com o vídeo de controle, notou-se o "efeito do céu" na estimação de Fluxo Ótico. Pontos muito distantes da câmera apresentam movimento relativo muito pequeno ou inexistente. O mesmo acontece com superfícies homogêneas, uma vez que o fluxo ótico estimado é denso e não baseado em pontos de interesse. Eventualmente, todo ponto em superficies homogêneas e de "céu" serão tratados como objetos e isso por si só

Tabela 2: Resultados para o movimento translacional puro.

X_{FOE} / Y_{FOE}	-1000	-700	-400	-100	200	500	800
-1000	0,6348	0,5811	0,5811	0,6369	0,7355	0,244	0,2449
-700	1,0000	0,6275	0,5811	0,5978	0,7000	0,2449	0,2449
-400	1,0000	1,0000	1,0000	0,5811	0,6681	0,2449	0,9891
-100	1,0000	0,9724	1,0000	1,0000	0,989	0,9891	0,989
200	1,0000	1,0000	1,0000	1,0000	0,9318	0,955	0,9550
500	1,0000	1,0000	1,0000	1,0000	0,9202	0,9550	0,9550
800	1,0000	1,0000	1,0000	0,06159	0,2942	0,1202	0,9550

Tabela 3: Resultados para o movimento afim.

a1	a2	a3	b1	b2	b3	Taxa de acerto (%)
0,04	0,08	-5	0	0,1	-10	70,362
0,04	0,08	-5	0,01	0	-10	24,493
0,04	0,08	-5	0,01	0,01	-10	73,333
0,04	0,08	-5	0,01	0,02	-10	77,899
0,04	0,08	-5	0,01	0,03	-10	73,261
0,04	0,08	-5	0,01	0,04	-10	69,203
0,04	0,08	-5	0,01	0,05	-10	67,826
0,04	0,08	-5	0,01	0,06	-10	67,609
0,04	0,08	-5	0,01	0,07	-10	66,377
0,04	0,08	-5	0,01	0,08	-10	66,087
0,04	0,08	-5	0,01	0,09	-10	65,942
0,04	0,08	-5	0,01	0,1	-10	65,652
0,04	0,08	-5	0,02	0	-10	52,029

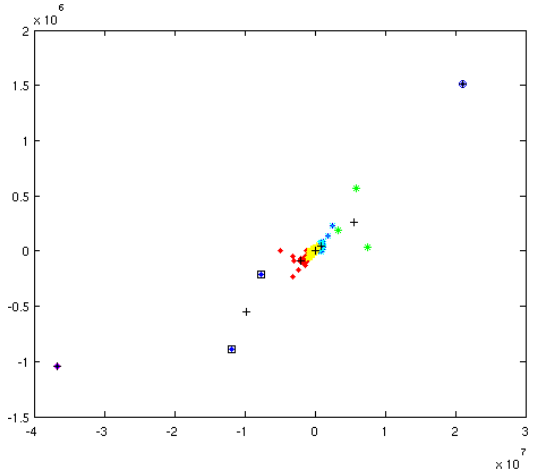

Figura 33: Dispersão espacial dos Focos de Expansão.
Fonte: Autoria própria, adaptação de Kuiaski et al. (2011).

traduz-se na inclusão de erros de estimação.

Uma abordagem menos ingênua é considerar a localização espacial dos focos de expansão relativos e agrupá-los. A primeira ideia foi utilizar um algoritmo de agrupamento, por *K-Means*. A figura 33 mostra a disposição espacial dos focos de expansão relativos para o vídeo de controle. Nesse caso, os resultados foram agrupados em sete conjuntos. Essa quantidade foi escolhida empiricamente e o Foco de Expansão considerado foi encontrado como sendo o centróide da região com maior densidade de focos de expansão relativos. Existem dois cenários estudados: Sem filtrar o céu e regiões homogêneas, a taxa de erro de classificação de objeto chega a patamares bastante elevados de 80% e de fundo a 5,9%; Quando o efeito céu é tratado, essa taxa de erro cai a 30% para objetos e a 5% para fundo. O interessante dessa abordagem sobre o vídeo de controle, *car.avi* é que ele possui uma pequena componente de rotação.

Esse procedimento, realizado sobre a base de dados sintética resultou em uma taxa de acerto de 79% para movimento translacional e de 64% para movimento afim. O que pôde ser notado é que o erro mais comum desse tipo de classificação é estimar objeto como fundo, dependendo da margem de erro considerada e do movimento do objeto. Se o movimento do objeto não for rígido ou se a sua velocidade for muito pequena, erros de estimação podem acarretar na sua classificação como fundo.

7 CONCLUSÃO

A hipótese inicial de que a Análise de Componentes Independentes seria uma técnica adequada para segmentação das componentes de movimento em um campo de Fluxo Ótico não foi confirmada. A segmentação de movimento por si só é um problema ainda mal resolvido do processamento de vídeo. Quando falamos em estimação de Fluxo Ótico, falamos intrinsecamente de erros de estimação, aproximações e hipóteses. Ao juntarmos essas características com uma técnica estatística como a estimação ICA, podemos esperar que muitos processos intermediários precisam ser realizados para que algum resultado consiga ser consistente, como transformações de domínio.

Isso é um bom indicativo de que ICA por si só pode não ser suficiente para se trabalhar com vídeo ou Fluxo Ótico, da forma como ele é estimado, através da representação vetorial do movimento. Os indícios dados por Bell e Sejnowski (1997) e por Gibson (1966) levam à conclusão de que, de alguma forma, ICA é um processo natural de retirada de redundâncias e que deve existir um modelo espaço-temporal que se aplique para se separar o *egomotion* dos demais movimentos que são detetados na retina. Isso leva a crer que não é a estimação ICA que não se encaixa ao Fluxo Ótico, mas o Fluxo Ótico, da forma como é estimado atualmente, que não se encaixa à estimação ICA. Se a Teoria Ecológica estiver correta, existe um novo paradigma de captação e percepção de movimento que não pode ser reproduzido computacionalmente ainda. Esse caráter ecológico é, talvez, o elo que falta para que um trabalho na linha desta pesquisa seja conclusivo. Uma análise dessa natureza foge ao escopo deste trabalho e demandaria mais tempo para ser avaliado, ficando como questão para trabalhos futuros.

Observou-se, ainda, que as restrições impostas à estimadores, como o de Horn e Schunck, podem induzir a uma distribuição gaussiana, como forma de resolução do problema da abertura. Nesse caso, tais restrições por si só invalidam a utilização dos modelos de ICA.

Apesar de não ter alcançado plenamente o objetivo de segmentar movimento, esta pesquisa foi válida para mostrar que ao mesmo tempo que ICA é uma ferramenta poderosa, também é limitada e inexplorada. Da mesma forma que pode ser substituída por técnicas al-

ternativas que eventualmente são menos custosas computacionalmente. Uma contribuição foi a criação de uma base de dados que contemple a classificação dos vetores de movimento e para atrair a atenção para esta linha de pesquisa, que pode gerar frutos em trabalhos futuros.

Existem algumas falhas tanto na primeira parte, de estimação de ICA, quanto na segunda, de técnicas alternativas por foco de expansão. Quando foram iniciados os experimentos de ICA sobre campos de Fluxo Ótico estimados ainda não se havia cogitado a necessidade de criar uma base de dados com gabarito. Esse diferencial só foi percebido quando os primeiros algoritmos foram rodados e não havia como mensurar os acertos e erros. Nessa mesma época, foi observada a inconveniência das limitações de ICA. A necessidade de se trabalhar independentemente com as componentes X e Y do deslocamento implicava em um pós processamento para a escolha de qual componente independente relativa a X seria correspondente a qual componente indepentende relativa a Y. Da mesma forma, não era garantido que a perda de energia quando se aplicava ICA às componentes de movimento relativas a X era a mesma perda relativa ao trabalho em Y. Isso por si só é motivo para inviabilizar tal técnica, pois se as componentes X e Y não foram escalonadas pelo mesmo fator, o ângulo do vetor de movimento se altera e pode diferir do modelo de movimento do fundo.

Outro ponto que alterou o rumo desta pesquisa foi a impossibilidade de replicar o experimento deBell e Sejnowski (1997). Várias tentativas foram feitas utilizando a própria biblioteca IT++, que está implementada em C++ e diz-se de alto desempenho. Entretanto, o tempo de processamento para mais do que seis componentes independentes mostrou-se impraticável. Com as 144 variáveis aleatórias utilizadas pelos autores, não havia capacidade computacional necessária. Esse talvez foi um dos indícios mais fortes de que contornar todas essas limitações tecnológicas e de paradigmas seria um projeto maior do que este trabalho se propunha.

A ideia de utilização de Focos de Expansão surgiu como uma forma rápida e simplificada de alcançar o mesmo objetivo. O objetivo era trabalhar inicialmente com translação pura, que constitui o caso mais simples, e, se possível, estender para outros modelos de movimento. Uma análise do quanto a variação no valor dos parâmetros interfere no resultado final poderia ter sido realizada. Entretanto, a ideia inicial era apenas provar a validade dessas técnicas como classificadores e segmentadores dos vetores de movimento. Esse trabalho ocorreu em duas partes distintas e talvez por isso tenha tomado mais tempo do que o esperado. Uma, da implementação das lógicas de estimação de Fluxo Ótico e de cálculo dos Focos de Expansão relativos para cada par de vetores de movimento e outra do agrupamento desses Focos de Expansão relativos calculados e a estimação do Foco de Expansão de *egomotion* a ser usado como classificador. Essa segunda parte foi realizada utilizando-se o Matlab ®.

A maior dificuldade para desenvolver esta pesquisa foi a falta de linhas de pesquisa na área. No universo do processamento de vídeo e mesmo no de imagens, ICA ainda é um mundo inexplorado, cujas aplicações ainda são muito restritas. Na estimação de movimento por Foco de Expansão, a falta de uma base de dados e a pouca literatura para o caso de uma única câmera sem informações *a priori* foi o maior entrave.

7.1 TRABALHOS FUTUROS

Este trabalho mostra duas possíveis linhas de pesquisa futuras. A primeira, dentro do contexto de ICA e da revisão do paradigma de processamento de vídeo. A Teoria Ecológica de Gibson (1979) sugere uma maior integração do sistema de captação das câmeras com o software de processamento, talvez como um pré-processamento a nível de *hardware*. Ainda nesse contexto, uma outra abordagem parte da definição dos Tensores Estruturais de Farnebäck (2002) e a definição de ICA por métodos tensoriais e cumulantes presentes em Hyvärinen et al. (2001). Uma primeira análise indica a possibilidade de se juntarem esses conhecimentos sob uma perspectiva mais coerente e funcional.

A segunda envolve a estimação parametrizada de movimento. O agrupamento realizado neste trabalho utilizou um número fixo e empírico de grupos via *K-Means*. Entretanto, uma alternativa mais robusta envolveria a utilização de uma rede neural artificial do tipo GNG (*Growing Neural Gas*) (FRITZKE, 1995) ou do tipo GWR (*Grow When Required*) (MARSLAND et al., 2002). A utilização de um algoritmo de otimização, como um algoritmo genético, também seria uma opção para a estimação dos parâmetros da modelagem de movimento. Também, uma análise qualitativa desses modelos paramétricos de movimento dentro de vários cenários seria um bom ponto de partida para qualquer pesquisa na área.

REFERÊNCIAS

BAKER, S.; MATTHEWS, I. Lucas-Kanade 20 years on: A unifying framework. **International Journal of Computer Vision**, p. 221–255, 2004.

BARLOW, H. Unsupervised learning. **Neural Computation**, v. 1, p. 295–311, 1989.

BAY, H. et al. Speeded-up robust features (surf). **Computer Vision and Image Understanding**, v. 10, p. 346–359, 2008.

BELL, A.; SEJNOWSKI, T. The 'Independent Components' of natural scenes are edge filters. **Vision Research.**, p. 3327–3338, 1997.

BISHOP, C. **Pattern recognition and Machine Learning.** [S.l.]: Springer, 2006.

BRUSS, A.; HORN, B. Passive navigation. **Computer Graphics and Image Processing**, v. 21, p. 3–20, 1983.

CAM, L. L. The central limit theorem around 1935. **Statistical Science**, v. 1, p. 78–91, 1986.

CARDOSO, J. Source separation using higher order moments. **In Proc. ICASSP'89.**, p. 2109–2112, 1989.

CARDOSO, J.; LAHELD, B. Equivariant adaptive source separation. **IEEE Trans. on Signal Processing**, v. 44, p. 3017–3030, 1996.

COMON, P. Independent component analysis - a new concept? **Signal Processing**, v. 36, p. 287–314, 1994.

FARNEBÄCK, G. Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field. **In Proceedings of the 8th IEEE International Conference on Computer Vision**, 2001.

FARNEBÄCK, G. **Polynomial Expansion for Orientation and Motion Estimation**. Tese (Doutorado) — Linköping University, Sweden, 2002.

FRITZKE, B. A growing neural gas network learns topologies. In: **Advances in Neural Information Processing Systems 7**. [S.l.]: MIT Press, 1995. p. 625–632.

GIBSON, J. The theory of affordances. **In Perceiving, Acting and Knowing, Eds. Robert Shaw and John Bransford**, 1977.

GIBSON, J. J. **The Perception of the Visual World**. Boston: Houghton Mifflin, 1950.

GIBSON, J. J. **The Senses Considered as Perceptual Systems**. Boston: Houghton Mifflin, 1966.

GIBSON, J. J. **The Ecological Approach to Visual Perception**. Boston: Houghton Mifflin, 1979.

HARRIS, C.; STEPHENS, M. A combined corner and edge detector. **In Proceedings of the 4th Alvey Vision Conference**, p. 147–151, 1988.

HEEGER, D.; JEPSON, A. Subspace methods for recovering rigid motion. I. algorithm and implementation. **International Journal of Computer Vision**, v. 7, p. 95–117, 1992.

HERAULT, J.; JUTTEN, C. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. **Signal Processing.**, p. 1–10, 1991.

HERAULT, J.; JUTTEN, C.; ANS, B. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. **Actes du Xème colloque GRETSI**, p. 1017–1022, 1985.

HORN, B.; SCHUNCK, B. Determining optical flow. **Artificial Iintelligence**, 1980.

HUBER, D. **The Computer Vision home Page**. 2004. Disponível em: <http://www.cs.cmu.edu/Groups/cil/vision.html>.

HYVÄRINEN, A.; KARHUNEN, J.; OJA, E. **Independent Component Analysis**. [S.l.]: John Wiley and Sons, inc., 2001.

KANATANI, K. 3-d interpretation of optical flow by renormalization. **International Journal of Computer Vision**, v. 11, p. 267–282, 1993.

KUIASKI, J.; LAZARETTI, A.; NETO, H. V. Focus of expansion for motion segmentation from a single camera. **Anais do VII Workshop de Visão Computacional (WVC 2011)**, 2011.

LACOUME, J. e. a. Blind separation of wide-band sources in the frequency domain. **In Proceedings of ICASSP-95**, v. 3, p. 2080–2083, 1995.

LONGUET-HIGGINS, H.; PRAZDNY, K. The interpretation of a moving retinal image. **Proceedings of the Royal Society of London**, p. 385–397, 1980.

LUCAS, B.; KANADE, T. An iterative image registration technique with an application to stereo vision. **In Proceedings of Imaging Understanding Workshop**, p. 121–130, 1981.

MARR, D. **Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.** [S.l.]: V. H. Freeman, 1982.

MARSLAND, S.; SHAPIRO, J.; NEHMZOW, U. A self-organizing network that grows when required. **Neural Networks**, v. 15, p. 1041–1058, 2002.

MERCK. **Manual Merck de Informação Médica**. 2010. Disponível em: <http://mmspf.msdonline.com.br/pacientes/manual˙merck/secao˙20/cap˙227.html>.

MORAVEC, H. Obstacle avoidance and navigation in the real world by a seeing robot rover. **Tech Report CMU-RI-TR-3 Carnegie-Mellon University.**, 1980.

NETO, V. O.; GOMES, D. Comparação de métodos para localização de fluxo ótico em sequências de imagens. **Seminário de Projeto e Análise de Algoritmos.**, 2011.

OJA, E.; HYVARINEN, A. A fast fixed-point algorithm for independent component analysis. **Neural Computation**, v. 9, p. 1483–1492, 1997.

OLSHAUSEN, A.; FIELD, D. Natural image statistics and efficient coding. **Network.**, v. 7, p. 333–340, 1996.

OPENCV. **OpenCV Wiki**. 2012. Disponível em: <http://opencv.willowgarage.com/wiki/>.

PALMER, S. **Vision Science: Photons to Phenomenology**. Cambridge, MA: Bradford Books, 1999.

PAPOULIS, A. **Probability, random variables, stochastic processes.** [S.l.]: MGH, 1991.

PRAZDNY, K. Egomotion and relative depth map from optical flow. **Biological Cybernetics**, v. 36, p. 87–102, 1980.

REED, E. S. **James J. Gibson and the Psychology of Perception**. [S.l.]: Yale University Press, 1989.

RETINA, G. **Olho: como é a sua estrutura.** 2009. Disponível em: <http://www.gruporetina.org.br/olho.html>.

SHI, J.; TOMASI, C. Good features to track. **In Proceedings of the 9th IEEE Conference on Computer Vision and Pattern Recognition. Springer.**, 1994.

STAVENS, D. **David Stavens' Homepage**. 2005. Disponível em: <http://www.cs.stanford.edu/people/dstavens/cs223b/>.

STILLER, C.; KONRAD, J. Estimating motion in image sequences: A tutorial on modeling and computation of 2d motion. **IEEE Signal Process. Mag.**, v. 6, p. 70–91, 1999.

TOMASI, C.; SHI, J. Direction of heading from image deformations. **In Proceedings of IEEE Computer Vision and Pattern Recognition**, p. 422–427, 1993.

TRUCCO, E.; VERRI, A. **Introductory Techniques for 3-D Computer Vision.** [S.l.]: Prentice-Hall, 1998.

ULLMAN, S. **The Interpretation of Visual Motion**. Cambridge, MA: MIT Press, 1979.

WEITHEIMER, M. Laws of organization in perceptual forms. **Psycologische Forschung**, v. 4, p. 301–350, 1923.