

**UNIVERSIDADE DE CAXIAS DO SUL
ÁREA DO CONHECIMENTO DE CIÊNCIAS EXATAS E ENGENHARIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA - PPGMEC**

GUSTAVO DE CAMARGO

**REGISTRO DE IMAGENS APLICADO À MANIPULAÇÃO INDUSTRIAL
ROBOTIZADA**

**CAXIAS DO SUL
2022**

GUSTAVO DE CAMARGO

**REGISTRO DE IMAGENS APLICADO À MANIPULAÇÃO INDUSTRIAL
ROBOTIZADA**

Dissertação submetida ao Corpo Docente do Programa de Pós-Graduação em Engenharia Mecânica, PPGMEC, da Universidade de Caxias do Sul, como parte dos requisitos necessários para a obtenção do Título de Mestre em Engenharia Mecânica.
Área de concentração: Projeto e Fabricação.

Orientador:
Prof. Dr. Guilherme Holsbach Costa

**CAXIAS DO SUL
2022**

Dados Internacionais de Catalogação na Publicação (CIP)
Universidade de Caxias do Sul
Sistema de Bibliotecas UCS - Processamento Técnico

C172r Camargo, Gustavo de
Registro de imagens aplicado à manipulação industrial robotizada
[recurso eletrônico] / Gustavo de Camargo. – 2022.
Dados eletrônicos.
Dissertação (Mestrado) - Universidade de Caxias do Sul, Programa de
Pós-Graduação em Engenharia Mecânica, 2022.
Orientação: Guilherme Holsbach Costa.
Modo de acesso: World Wide Web
Disponível em: <https://repositorio.ucs.br>
1. Registro de imagem. 2. Robótica. I. Costa, Guilherme Holsbach,
orient. II. Título.

CDU 2. ed.: 004.896

Catalogação na fonte elaborada pela(o) bibliotecária(o)
Márcia Servi Gonçalves - CRB 10/1500

GUSTAVO DE CAMARGO

**REGISTRO DE IMAGENS APLICADO À MANIPULAÇÃO INDUSTRIAL
ROBOTIZADA**

Dissertação submetida ao Corpo Docente do Programa de Pós-Graduação em Engenharia Mecânica, PPGMEC, da Universidade de Caxias do Sul, como parte dos requisitos necessários para a obtenção do Título de Mestre em Engenharia Mecânica.
Área de concentração: Projeto e Fabricação.

Orientador:
Prof. Dr. Guilherme Holsbach Costa

Aprovado em ____/____/____

Banca Examinadora

Prof. Dr. Guilherme Holsbach Costa
Universidade de Caxias do Sul - UCS

Prof^a. Dr^a. Adriana Miorelli Adami
Universidade de Caxias do Sul - UCS

Prof^a. Dr^a. Marilda Machado Spindola
Universidade de Caxias do Sul - UCS

Prof. Dr. Ricardo Augusto Borsoi
Université de Lorraine (France)

*"As I review the events of my past life I realize how subtle
are the influences that shape our destinies"*
Nikola Tesla

RESUMO

O referenciamento de robôs em aplicações industriais de manipulação de objetos é tipicamente baseado em técnicas de reconhecimento de padrões. Entretanto, esse tipo de solução é fortemente dependente do treinamento, é computacionalmente custosa e geralmente não é passível de análise teórica de desempenho. Neste trabalho, é proposta a aplicação de algoritmos de registro de imagens para estimar a posição e orientação dos objetos a serem manipulados. Para tanto, propõe-se o uso de um método baseado em Correlação de Fase, de forma a segmentar o objeto de interesse e, simultaneamente, estimar os parâmetros de interesse. Como uma das principais contribuições, são apresentadas considerações sobre o registro de imagens por Correlação de Fase, destacando-se limitações que aparentemente passaram despercebidas pela comunidade científica nos últimos 25 anos. Com base nessas considerações, sob condições de teste de interesse da aplicação e assumindo-se o erro absoluto médio como figura de mérito, o método proposto melhora o desempenho de estimação de movimentos de rotação em mais de 1000% em relação à forma em que o método de Correlação de Fase se apresenta na literatura.

Palavras-chave: Registro de imagens. Correlação de Fase. Manipulação robótica.

ABSTRACT

Robot referencing in industrial object manipulation applications is typically based on pattern recognition techniques. However, this type of solution is strongly dependent on training, is computationally expensive, and is generally not susceptible to a theoretical performance analysis. In this work, the application of image registration algorithms is proposed to estimate the position and orientation of the objects to be manipulated. Therefore, we propose the use of a method based on Phase Correlation, in order to segment the object of interest and, simultaneously, estimate the parameters of interest. As one of the main contributions, considerations about the registration of images using Phase Correlation are presented, highlighting limitations that apparently went unnoticed by the scientific community in the last 25 years. Based on these considerations, under test conditions of interest to the application and assuming the mean absolute error as a figure of merit, the proposed method improves the performance of estimation of rotational movements by more than 1000% in relation to the way in which the Phase Correlation method is presented in the literature.

Keywords: Image registration. Phase Correlation. Robotic manipulation.

LISTA DE FIGURAS

Figura 1 – Etapas do registro de imagem	16
Figura 2 – Magnitude da resposta em frequência do filtro passa-baixas	26
Figura 3 – Magnitude da resposta em frequência do filtro passa-altas	26
Figura 4 – Exemplo de transformação do plano cartesiano para o plano polar por meio de <i>mapeamento inverso</i>	31
Figura 5 – Exemplo de efeito da rotação no plano cartesiano sobre o plano polar	32
Figura 6 – Exemplo de transformação do plano cartesiano para o plano polar por meio de <i>mapeamento direto</i>	33
Figura 7 – Exemplos de imagens de referência e de entrada usadas no teste I (objetos iguais sobre fundo preto).	34
Figura 8 – Exemplos de imagens de referência e de entrada usadas no teste II (objetos iguais sobre fundos distintos).	35
Figura 9 – Exemplos de imagens de referência e de entrada usadas no teste III (objetos distintos sobre fundo preto).	36
Figura 10 – Exemplos de imagens de referência e de entrada usadas no teste IV (objetos distintos sobre fundos distintos).	36
Figura 11 – Exemplos de imagens de referência e de entrada usadas no teste V (imagens naturais).	37
Figura 12 – Resultados do registro de imagens utilizando o método original e o método proposto	40
Figura 13 – Resultado da estimação de rotação pelo método original	41
Figura 14 – Exemplo de Mapeamento	50
Figura 15 – Mapeamento de Cartesiano para Log-Polar	51

LISTA DE SIGLAS

CC	Correlação Cruzada
P,D&I	Pesquisa, Desenvolvimento e Inovação
CDI	Centro de Desenvolvimento e Inovação
CPs	<i>Control Points</i>
SIFT	<i>Scale-Invariant Feature Transform</i>
DTFT	<i>Discrete Time Fourier Transform</i>
FFT	<i>Fast Fourier Transform</i>
MPFFT	<i>Multilayer-Pseudopolar Fractional Fourier Transform</i>
PPFFT	<i>Pseudo-Polar Fast Fourier Transform</i>
MLFFT	<i>Multilayer Fast Fourier Transform</i>

SUMÁRIO

1	INTRODUÇÃO	10
1.1	MOTIVAÇÃO	11
1.2	JUSTIFICATIVA E ENCAMINHAMENTO	12
1.3	OBJETIVOS	14
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	O REGISTRO DE IMAGENS	15
2.2	ABORDAGENS E MÉTODOS DE REGISTRO DE IMAGEM	16
2.3	NOTAÇÃO	19
2.4	REGISTRO BASEADO NA TRANSFORMADA DE FOURIER	20
2.4.1	Estimação de Movimentos com Translação e Rotação	21
2.4.2	Estimação de Movimentos de Reescala	22
2.4.3	Estimação de Movimentos com Translação, Rotação e Reescala	23
2.5	QUESTÕES ACERCA DA IMPLEMENTAÇÃO DO MÉTODO	25
3	MÉTODO PROPOSTO	30
4	AVALIAÇÃO DO MÉTODO PROPOSTO	34
4.1	ESTIMAÇÃO DE MOVIMENTOS DE TRANSLAÇÃO	38
4.2	ESTIMAÇÃO DE MOVIMENTOS DE ROTAÇÃO	38
4.3	RESULTADOS PARA MOVIMENTOS DE ROTAÇÃO E TRANSLAÇÃO	39
5	CONCLUSÃO	42
	REFERÊNCIAS	44
	APÊNDICE A – ANÁLISE DE CORRELAÇÃO E DENSIDADE ESPECTRAL	49
	APÊNDICE B – MAPEAMENTO LOG-POLAR	50

1 INTRODUÇÃO

Sistemas de visão computacional são sistemas computacionais dotados de câmeras de vídeo, que se baseiam no processamento de imagens para a tomada de decisões em processos automatizados. A visão computacional pode ser utilizada para a supervisão de processos, controle de qualidade, realimentação de malhas de controle, entre outras finalidades. Dentre as aplicações da visão computacional, a manipulação robótica de objetos é demandada em vários setores da indústria e, portanto, de especial interesse do ponto de vista de inovação. Quando os objetos são apresentados ao manipulador de forma organizada e em posições conhecidas, os sistemas de visão são dispensados. Entretanto, em muitos casos, o robô precisa identificar a posição e a orientação espacial do objeto a ser manipulado. Para isso, primeiro é necessário detectar a presença do objeto de interesse na cena, o que em muitas abordagens já resulta na identificação da posição. Como será discutido a seguir, essa *detecção* é geralmente feita por (ou em conjunto com) técnicas de *reconhecimento de padrões*, que neste trabalho é tratado como sinônimo de *classificação de padrões* ou de objetos.

A detecção de objetos tem sido amplamente estudada na área de visão computacional (KIM; RO, 2019). Esse tipo de técnica visa identificar a presença e geralmente a posição de um determinado objeto na cena, porém geralmente não abrange a identificação da orientação do objeto detectado, como discutido por Cheraghi e Sheikh (2012) e por Wang et al. (2018a). A detecção de objetos, portanto, dá conta apenas de parte da solução investigada neste trabalho.

Identificar o ângulo de rotação em que um determinado objeto se apresenta implica na necessidade de uso de um segundo classificador. Para evitar o uso de um classificador com, por exemplo, 360 classes, se desejada uma resolução de um grau¹, é comum a aplicação de redes de regressão (WANG et al., 2018b; BAO; YE, 2016). Entretanto, essas técnicas dependem de uma etapa de treinamento, que demanda a construção de uma base de dados que seja estatisticamente representativa (o que é um conceito de difícil avaliação, neste caso) do universo de imagens que podem ser apresentadas ao sistema de visão. Isso torna o processo de treinamento subjetivo e custoso em termos de tempo. Portanto, ainda que o referenciamento de robôs em aplicações de manipulação de objetos seja geralmente direcionado a técnicas de reconhecimento de padrões (ANTIPOV et al., 2017; JAHANANDISH; FEY; HOYT, 2019; ZHU; LEI; TIAN, 2008; BRUCE; VELOSO, 2003), as soluções encontradas na literatura potencialmente não atingem, em ambiente não controlado, níveis de acurácia na ordem de parte por milhão, comumente demandados em linhas de produção (SCHRUNNER et al., 2018; NGUYEN et al., 2018; OMAROV; CHO, 2017; KHARE, 2015).

Tratando especificamente de detecção de objetos em movimento, é comum que os métodos

¹A classificação entre um número muito grande de classes, com diferenças substancialmente sutis (no caso, diferença de um grau na orientação de um objeto cuja imagem é muitas vezes adquirida em ambiente não controlado, com variações de iluminação, por exemplo), pode recair em taxas de erro substancialmente altas para a aplicação em questão (RISTIN et al., 2014; NILSBACK; ZISSERMAN, 2008; MAI et al., 2015; JOSHI; PORIKLI; PAPANIKOLOPOULOS, 2009; FUKUNAGA; FLICK, 1984)

explorem a diferença entre quadros consecutivos de uma sequência de vídeo ou entre cada quadro e uma imagem padrão, previamente adquirida, que contenha apenas o fundo da cena (*background*) (FEFILATYEV; GOLDFOF; LEMBKE, 2010; SCHUBERT; MIKOLAJCZYK, 2014). Regiões com pequenas diferenças indicam ausência de movimento, no primeiro caso, ou que a imagem sob análise contém apenas o fundo esperado, no segundo caso, enquanto regiões com diferenças mais expressivas indicam a presença de movimento ou a presença de um objeto de interesse da detecção (*foreground*). Cheraghi e Sheikh (2012) propõem um método com essa base, porém dedicado a aquisições com câmeras em movimento. Para tanto, é realizado o *registro* — técnica de transformação geométrica que visa alinhar duas imagens, conforme será discutido na Seção 2.1— entre as imagens, de forma que as cenas fiquem alinhadas e possam ser comparadas propiciando a segmentação o objeto de interesse a partir da diferença entre elas. Entretanto, ainda que fosse considerado para a aplicação de manipulação robótica, o método proposto por Cheraghi e Sheikh (2012) igualmente não seria capaz de tratar adequadamente movimentos de rotação, por exemplo. Mais que isso, em muitas aplicações industriais os objetos de interesses são transportados por esteiras e as câmeras são estáticas em relação a essas esteiras. Sendo assim, registrar as imagens com base no fundo da cena não se justifica, visto que não há deslocamento relativo entre o fundo em diferentes aquisições.

Motivado pelo trabalho de Cheraghi e Sheikh (2012), neste trabalho é proposto o uso de algoritmos de registro de imagens aplicados não ao fundo da cena, mas diretamente ao objeto de interesse. Da função de transformação geométrica estimada no processo de registro é possível extrair os parâmetros de localização e de orientação de interesse deste trabalho. Dependendo da abordagem que os originam, os algoritmos de registro podem ser robustos a variações de iluminação e à presença de inovações² (FENG; YI-CAN, 2019; JIANG et al., 2021). Essa capacidade de robustez pode tornar a abordagem proposta uma candidata a substituir os métodos de reconhecimento de padrões, nessa aplicação, evitando a necessidade da composição de conjuntos de treinamento (baseando-se em um único padrão).

1.1 MOTIVAÇÃO

Este trabalho foi fomentado pela Máquinas Sanmartin LTDA, fundada em 1948 na cidade de Mercedes, na Argentina. Essa empresa possui filiais em Caxias do Sul - RS, no Brasil, aberta em 1976 e no México, aberta em 1979. É líder na América Latina e uma das principais empresas do mundo na produção de máquinas dedicadas ao transporte, empacotamento e paletização/despaletização de bebidas. No ano de 2019 a Sanmartin inaugurou seu Centro de Desenvolvimento e Inovação, o CDI Sanmartin, também na cidade de Caxias do Sul, sendo ele um espaço com ampla área e infraestrutura para P,D&I, onde parte desta pesquisa foi desenvolvida e direcionada.

²São usualmente chamadas de inovações (ou *outliers*) os dados que não correspondem às hipóteses (estatísticas) assumidas na proposição da solução.

Um diagnóstico realizado dentro do CDI identificou que um sistema de visão capaz de detectar objetos e reconhecer sua posição e orientação em uma cena, sob condições particulares de movimento, seria capaz de atender a diversas aplicações de interesse da empresa, na direção do desenvolvimento de novos produtos. A aplicação que motiva o presente trabalho se trata, então, do referenciamento de manipuladores industriais equipados com sistemas de visão computacional, como, por exemplo, manipuladores para seleção de objetos em esteiras transportadoras. Para isso, será assumida a hipótese de movimentos globais³, translacionais e de rotação no plano da imagem (em torno de um eixo paralelo ao eixo focal da lente do sistema de aquisição de imagens). É ainda desejável que a solução possa ser expandida para movimentos que contenham variações de escala (*zoom*). A verificação dessa hipótese depende ainda de que o objeto a ser manipulado seja o único presente na cena, diferentemente de um sistema de *bin picking*⁴, por exemplo. Havendo outros detalhes presentes na imagem, como estruturas mecânicas da esteira transportadora, uma etapa de pré-processamento pode ser realizada de forma a segmentar o objeto de interesse, etapa esta que também será abordada neste trabalho.

1.2 JUSTIFICATIVA E ENCAMINHAMENTO

Em face à aplicação que motiva este trabalho e à contextualização feita no início desta introdução, explora-se o uso de técnicas de registro na solução almejada. Os algoritmos de registro geralmente são propostos a partir de uma hipótese sobre a natureza do movimento presente na cena (SHARMA; GOYAL, 2013). Por exemplo, algoritmos dedicados ao registro de imagens com movimentos translacionais e globais podem se restringir à estimação de apenas dois parâmetros; deslocamento na direção das linhas e das colunas da imagem. Se o algoritmo pressupõe a possibilidade de um movimento de rotação no plano da imagem, adiciona-se a estimação de mais um parâmetro, o ângulo de rotação.

O modelo de movimento mais abrangente é, seguramente, o que se conhece por fluxo óptico (*optical flow*), em que o movimento de cada pixel pode ser independente dos demais (ZITOVÁ; FLUSSER, 2003). Isto resulta em dois parâmetros a serem estimados para cada pixel da imagem, o que geralmente resulta em centenas ou milhares de variáveis a serem estimadas para cada par de imagens a ser registrado. Ainda que a hipótese de movimentos mais genéricos possa parecer, em primeira análise, mais interessante para que se obtenha uma solução genérica, cada parâmetro a ser estimado constitui-se também como uma fonte de erro para a solução. Sendo assim, se, por exemplo, é sabido que não há rotação na cena, estimar um parâmetro que é, por definição, nulo, implica apenas em um erro numérico adicional.

A aplicação em questão pressupõe movimentos de translação e rotação. Além disso, são

³É chamado de global o movimento em que todos os *pixels* da imagem possuem o mesmo comportamento (vetor de deslocamento).

⁴*Bin Picking* é utilizado em sistemas de Robótica Guiada por Visão em que são selecionadas e extraídas peças amontoadas aleatoriamente, utilizando um sistema de visão para a localização e um sistema robótico para a extração.

requisitos desejáveis para as aplicações da Sanmartin um tempo de processamento abaixo de 500 milissegundos e uma taxa de erro inferior a uma parte por milhão (0,0001%). Um número substancialmente grande de algoritmos de registro de imagens, seguindo diferentes abordagens e aplicáveis (bem ou mal) a esse modelo de movimento, pode ser encontrado na literatura (BROWN, 1992; MAINTZ; VIERGEVER, 1998; NAG, 2017; SAXENA; SINGH, 2014; SHARMA; GOYAL, 2013; ZITOVÁ; FLUSSER, 2003). As avaliações de desempenho de cada proposta nem sempre são consistentes, de forma que possam ser adequadamente comparadas em termos de desempenho e principalmente em relação ao custo computacional, que não é abordado na maioria dos artigos da área. Sendo assim, uma comparação detalhada e exaustiva das soluções existentes demanda grande esforço de implementação e avaliação. De outra forma, a seleção de um método de registro pode ser feita com base na melhor adequação ao modelo de movimento, o que conduz à estimação no domínio da frequência (ZITOVÁ; FLUSSER, 2003). Parâmetros de deslocamento, rotação e *zoom* podem ser adequadamente estimados no domínio da frequência, fazendo-se uso da propriedade do deslocamento da Transformada de Fourier (ZITOVÁ; FLUSSER, 2003). Adicionalmente, o processamento em domínio transformado é tradicionalmente vantajoso, sob condições específicas, em termos de custo computacional, o que reforça a escolha. Ainda mais, esse tipo de abordagem tende a propiciar a proposição de modelos teóricos para o comportamento da solução, diferente das abordagens morfológicas, geralmente não lineares, ou baseadas em redes neurais, que acabam por demandar avaliações de Monte Carlo (ZITOVÁ; FLUSSER, 2003).

Zitová e Flusser (2003) realizaram uma pesquisa sobre os métodos de registro de imagem e descreveram o Método de Correlação de Fase como sendo inicialmente proposto por Kuglin e Hines (1975) para o registro de imagens que possuíssem somente movimento de translação. Em seguida, Castro e Morandi (1987) adicionaram uma extensão ao Método de Correlação de Fase para abranger também movimentos de rotação. Se as imagens também possuem diferença de escala (*zoom*), o registro pode ser feito utilizando a combinação de mapeamento log-polar com a correlação de fase (REDDY; CHATTERJI, 1996). A partir de então, a literatura se foca em modelos de movimento que contenham *zoom* (KELLER; AVERBUCH; ISRAELI, 2005; PAN; QIN; CHEN, 2009; LI et al., 2012). Entretanto, em 2003, Stone, Tao e McGuire (2003) voltam a focar em movimentos rotacionais em busca de melhorar o desempenho do registro por esse tipo de abordagem, apontando o *aliasing* dependente de rotação como um dos principais fatores que degradam o registro de imagem por correlação de fase. Apontam ainda que o método proposto por Reddy e Chatterji (1996) acarreta resultados falsos-positivos, com picos de correlação em 0° , e falsos-negativos, com valores de correlações menores que 0,3.

Dentro do escopo de movimentos de translação e de rotação, que correspondem à base que leva à estimação de variações de escala, Reddy e Chatterji (1996) pré-processam as imagens a serem registradas por um filtro passa-altas, alegando uma melhora de desempenho, sem apresentar uma justificativa para tanto. Stone, Tao e McGuire (2003) ignoram essa filtragem e, no lugar dela, aplicam um janelamento como pré-processamento, alegando reduzir as correlações

que resultam nos resultados falsos-positivos (estimando reiteradamente com erro o ângulo de rotação, como sendo 0°). Entretanto, neste trabalho é mostrado que o problema que gera tais falsos-positivos não advém do problema de espalhamento espectral, que seria, por hipótese, atenuado pela técnica de janelamento proposta por Stone, Tao e McGuire (2003) e que tampouco a filtragem passa-altas proposta por Reddy e Chatterji (1996) é a melhor opção para resolver o problema. Como principal contribuição, então, mostra-se que tal problema é determinístico, sendo resultado de um mapeamento de coordenadas previsto no método, o que, aparentemente, passou despercebido da comunidade científica nos últimos 25 anos.

No Capítulo 2 é apresentada a fundamentação teórica sobre registro de imagens, suas etapas e as principais abordagens encontradas na literatura, bem como a abordagem de registro no domínio da frequência e detalhes de sua implementação. No Capítulo 3, é apresentado o método proposto com suas respectivas contribuições. No Capítulo 4 são apresentados os resultados obtidos. Por fim, no Capítulo 5 é apresentada a conclusão do trabalho.

1.3 OBJETIVOS

Objetivo Geral

Selecionar, implementar e avaliar um algoritmo capaz de identificar a posição e a orientação de objetos em uma esteira transportadora disponíveis à manipulação robótica.

Objetivos Específicos

- Identificar uma abordagem adequada à identificação da posição e da orientação dos objetos, no contexto da aplicação de interesse;
- Selecionar um método, dentro da abordagem identificada como adequada;
- Implementar e avaliar a aplicabilidade do método selecionado;
- Adequar o método à aplicação de interesse;
- Avaliar o desempenho do método com as contribuições propostas.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo o conceito de registro de imagens é abordado com mais detalhes, na Seção 2.1, e uma revisão das diferentes abordagens de registro é apresentada na Seção 2.2. O resultado dessa revisão é o direcionamento do trabalho para o processamento no domínio da frequência. Em face disso, uma revisão sobre as propriedades da Transformada de Fourier pertinentes à solução abordada é apresentada na Seção 2.4, de forma a fundamentar o algoritmo discutido na Seção 2.5.

2.1 O REGISTRO DE IMAGENS

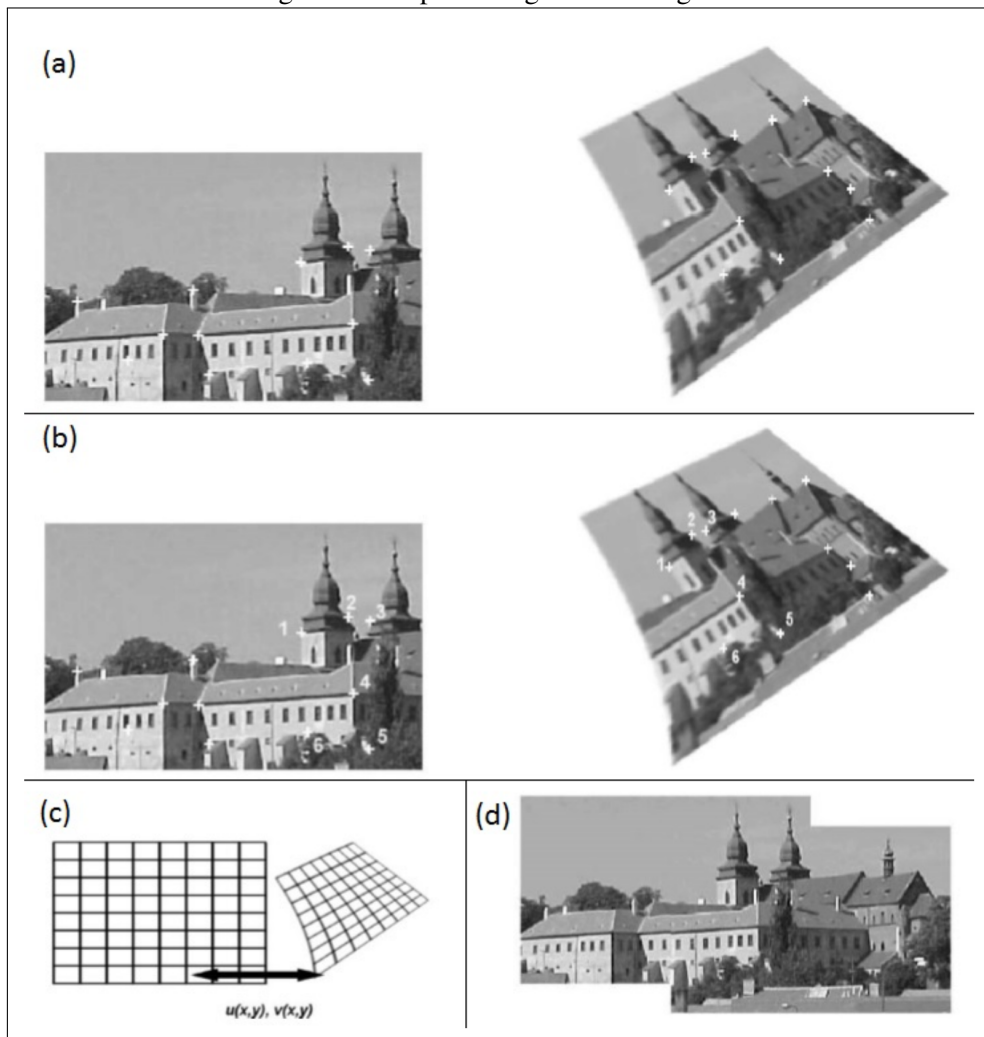
O Registro de imagens (*image registration*) é uma técnica de transformação geométrica que visa alinhar duas imagens. Esse tipo de técnica geralmente inclui a estimação do movimento relativo entre as duas cenas, ou seja, a estimação da função de transformação geométrica necessária para tal alinhamento (THEVENAZ; RUTTIMANN; UNSER, 1998).

Uma parte considerável dos métodos de registro de imagens pode ser dividida em quatro etapas: (a) detecção de características, (b) associação de características, (c) estimação do modelo da transformação e (d) re-amostragem e transformação da imagem. Ainda que essa divisão não se aplique a todos os métodos, além de ser representativa de uma grande classe de técnicas, ela é bastante didática no entendimento do processo de registro, conforme discutido a seguir.

- (a) Detecção das características — Características específicas de cada imagem, como objetos em realce, regiões de contorno fechado, arestas, interseções de linhas e cantos, são detectadas (de preferência) automaticamente. Essas características são descritas por coordenadas chamadas de pontos de controle (CPs — *Control Points*), ilustradas na Figura 1(a).
- (b) Associação das características — É estabelecida a correspondência entre as características detectadas na imagem de entrada e de referência. Descritores de características e medidas de similaridade, juntamente com relações espaciais entre as características, são usados para diferenciar uma característica de outra, relacionando cada ponto de controle presente nas duas imagens. Nem todo ponto de controle da imagem de entrada terá necessariamente um par correspondente na imagem de referência. Esse processo é ilustrado através dos rótulos (numéricos) adicionados na Figura 1(b).
- (c) Estimação dos parâmetros de transformação — Com base no modelo de movimento assumido pelo algoritmo, são estimados os parâmetros do mapeamento, que alinham a imagem de entrada com a imagem de referência, conforme representado na Figura 1(c). Os parâmetros das funções de mapeamento são calculados pela correspondência de características estabelecida no passo anterior.

(d) Re-amostragem e transformação da imagem — A imagem de entrada é transformada por meio das funções de mapeamento, aplicada a todos os *pixels*. Como resultado, as imagens podem ser espacialmente referenciadas, uma em relação à outra, como demonstra a Figura 1(d).

Figura 1 – Etapas do registro de imagem



(a) Detecção de características, (b) Associação de características, (c) Estimação dos parâmetros de transformação e (d) Re-amostragem e transformação da imagem. Fonte: Adaptado de (ZITOVÁ; FLUSSER, 2003)

2.2 ABORDAGENS E MÉTODOS DE REGISTRO DE IMAGEM

Diferentes autores têm classificado as técnicas de registro de imagens de diferentes formas (SAXENA; SINGH, 2014). Algumas dessas formas são apresentadas a seguir.

Brown (1992) classifica os métodos de registro de imagens em relação à forma de aquisição das imagens. Essa classificação resulta em quatro classes:

- Pontos de vista diferentes — a mesma imagem é capturada de pontos de vista diferentes.

Ainda que isto não esteja claro no trabalho de Brown (1992), nesta classe assume-se uma cena estática sendo capturada por um sensor de imagem em movimento;

- Tempos diferentes — a imagem é adquirida em instantes de tempo diferentes. Esta classe, então, se diferencia da anterior pela hipótese de que uma cena com objetos em movimento é adquirida por um sensor de imagens estático. Esta classe é geralmente considerada com o propósito de encontrar e mensurar mudanças que ocorreram entre dois instantes de tempo;
- Sensores diferentes — uma mesma cena, geralmente estática, é adquirida por diferentes sensores, geralmente ligeiramente desalinhados. Esta classe geralmente se aplica à fusão de sinais, como por exemplo em aplicações de sensoriamento remoto e em imagens médicas;
- Registro entre uma cena e um modelo (*Scene of model registration*) — o registro é feito entre uma imagem natural (adquirida por um sensor) e uma imagem sintética (simulada). O objetivo é localizar e comparar a imagem adquirida com um padrão sintético (modelo).

Zitová e Flusser (2003) separam os métodos em relação a duas possíveis abordagens de solução, resultando em duas classes:

- Métodos baseados em características — esta abordagem é baseada na extração de estruturas e características morfológicas das imagens, podendo elas serem entendidas como regiões, linhas e pontos;
- Métodos baseados em área — possuem ênfase na associação de características não morfológicas. Geralmente focam em características estatísticas ou em uma função de similaridade entre as imagens e recaem em um problema de otimização com solução recursiva. Como não há propriamente uma detecção de características nesta abordagem, então a primeira etapa do registro de imagem discutida na seção anterior é omitida.

Além da disposição do sistema de aquisição e das abordagens de solução, Maintz e Viergever (1998) destacam nove outros fatores que podem ser usados para classificar os métodos de registro. Dentre eles, os fatores pertinentes à aplicação de manipulação robótica são:

- Dimensionalidade — refere-se ao número de dimensões geométricas que qualquer espaço de imagem tem, podendo então ser o registro bidimensional ou tridimensional;
- Natureza da transformação — trata de hipóteses assumidas sobre o modelo dos movimentos presentes na cena. É dividida nas classes afim, projetiva, rígida ou curvada. Todavia, observe-se que, neste caso, pode ser considerada uma sobreposição das caracterizações propostas por Maintz e Viergever (1998) visto que, por exemplo, a transformação afim é seguramente uma transformação rígida.

- Domínio da transformação — classifica a transformação em local ou global. Movimentos globais são aqueles em que todos os *pixels* da imagem obedecem a uma mesma regra (função) de deslocamento. Movimentos locais são aqueles em que *pixels* de diferentes regiões da imagem seguem diferentes funções de deslocamento;
- Grau de interação — classifica o registro em relação ao grau de controle atribuído, ou demandado, ao ser humano pelo algoritmo. Alguns métodos demandam, por exemplo, que o usuário informe os pontos de controle, enquanto outros detectam esses pontos automaticamente;
- Procedimento de otimização — classifica os algoritmos em relação aos métodos iterativos utilizados para realizar numericamente o registro das imagens.

Uma das abordagens mais comuns para solucionar o problema de registro de imagens, baseada em características (ZITOVÁ; FLUSSER, 2003), é o uso de pontos de controle. Conhecendo-se a posição de algumas características em ambas imagens, chamadas de pontos de controle, é possível estimar um padrão de movimento para os demais pontos da imagem. Os pontos de controle podem ser selecionados nas imagens de forma manual ou automática (YANG; YUANLI, 2017; FONSECA; KENNEY, 1999; WANG et al., 2017; MENG; HAN; LI, 2017; HENEGHAN et al., 2002; PEI; JIA, 2018; FENG; YI-CAN, 2019). A principal desvantagem desse tipo de abordagem é que a detecção automática dos pontos de controle é fundamentalmente feita por técnicas morfológicas, fazendo com que as soluções finais sejam de difícil análise teórica de desempenho. Além disso, o custo computacional desse tipo de técnica pode ser alto quando consideradas as taxas de inspeção demandadas por algumas linhas de produção.

Um exemplo que tem se destacado na literatura é o SIFT (*Scale-Invariant Feature Transform*), com mais de 13 mil citações no Scopus (2020) e mais de 11 mil citações no WebOfScience (2020). Porém, Zhang, Zhang e Zeng (2011) apresentam tempos de processamento para o registro entre 2,4 e 9,2 segundos e Zhu et al. (2016) apresentam tempos de processamento, incluindo etapas de pré-processamento e de classificação, entre 13 e 414 segundos. Considerando a dinâmica das aplicações atuais da empresa Sanmartin, as quais necessitam tempos totais de processamento inferiores a 500 milissegundos, esse método não é, portanto, potencialmente adequado. Conforme discutido na Introdução, são potencialmente inadequados também os métodos que necessitam da composição de conjuntos de treinamento, como é o caso de algoritmos que utilizam técnicas de Inteligência Artificial.

Tong et al. (2019), assim como Zitová e Flusser (2003) concluíram que, pelo fato dos métodos baseados em área utilizarem diretamente as informações de intensidade para combinar áreas ou regiões, possuem vantagens em termos de precisão e de robustez. Além disso, métodos baseados em área não demandam uma pré-localização dos pontos de controle, o que pode reduzir o custo computacional.

Considerando as características de viabilidade de análise matemática, robustez e adequação à implementação em tempo-real descritas, este trabalho se direciona aos métodos baseados em

área. Sobre esses métodos, Zitová e Flusser (2003) destacam os métodos de correlação, os métodos baseados na transformada de Fourier e os métodos que utilizam informações mútuas.

Os métodos que utilizam correlação têm como representante clássico o método de correlação cruzada normalizada (do inglês, *normalized cross-correlation*) e suas modificações. Essa medida de similaridade é calculada a partir das imagens de entrada e de referência e seu valor máximo é pesquisado. Os pares para os quais o máximo é atingido são definidos como correspondentes. Se a precisão do registro do subpixel é exigida, a interpolação dos valores da medida de correlação precisa ser usada. Embora o registro baseado em correlação seja dedicado a imagens que se relacionem por um movimento de translação, ele também pode ser aplicado com sucesso a pares de imagens que se distinguam por sutis rotações de alterações de escala.

Os métodos baseados na Transformada de Fourier são usualmente considerados quando é necessária uma alta velocidade computacional (ZITOVÁ; FLUSSER, 2003). O Método de Correlação de Fase é baseado no Teorema do Deslocamento de Fourier (do inglês, *Fourier Shift Theorem*) e foi proposto originalmente para o registro de imagens transladadas. Como será discutido em detalhes na Seção 2.4, nele é calculado o espectro de potência cruzada das imagens e procurada a localização do pico de sua transformada inversa. Esse método supostamente apresenta forte robustez contra ruído correlacionado e dependente da frequência, além dos distúrbios de iluminação não uniformes.

Os métodos que utilizam informações mútuas apareceram recentemente e representam a principal técnica de registro multimodal, ou seja, de registro de imagens oriundas de sensores de diferentes naturezas (por exemplo, sensores para diferentes faixas espectrais). O registro de imagens multimodais não é uma tarefa trivial, pois busca uma correspondência entre imagens com diferentes características (ZITOVÁ; FLUSSER, 2003). Esse tipo de registro tem menor aplicação industrial, sendo entretanto largamente aplicado a imagens médicas e de sensoramento remoto (imagens hiper-espectrais).

Em suma, métodos baseados em área e com processamento no domínio da frequência tendem a ser preferidos para processamentos em tempo-real com altas taxas de aquisição e quando as condições de aquisição das imagens são variantes, como, por exemplo, variações na iluminação da cena (ZITOVÁ; FLUSSER, 2003). O registro de imagens realizado no domínio da frequência se origina com o Método de Correlação de Fase, baseado na Propriedade do Deslocamento da Transformada de Fourier. A fundamentação teórica desse método é apresentada na sessão seguinte.

2.3 NOTAÇÃO

Neste trabalho, variáveis matriciais serão denotadas por maiúsculas em negrito, a exemplo de \mathbf{X} . Vetores serão denotados por minúsculas em negrito, como em \mathbf{x} . Variáveis escalares serão denotadas por minúsculas em itálico, como em x . Escalares que representam constantes serão denotadas por maiúsculas em itálico, como em X . Por fim, variáveis representadas no

domínio da frequência serão sinalizadas com o diacrítico til, como em $\tilde{\mathbf{X}}$.

2.4 REGISTRO BASEADO NA TRANSFORMADA DE FOURIER

O Método de Correlação de Fase foi primeiramente descrito por Kuglin e Hines (1975). Foi utilizado para calcular somente translações entre duas imagens. Foi retomado por Reddy e Chatterji (1996) que adicionaram ao método a mensuração do ângulo de rotação e redimensionamento entre as imagens e descrevem o método como insensível à translação, rotação, redimensionamento e robusto ao ruído, além de possuir baixo custo computacional. Nesse trabalho, os autores determinam o deslocamento (na ocorrência de movimento translacional e global) entre duas imagens utilizando a propriedade do deslocamento da Transformada de Fourier. Para isso, considere-se duas imagens \mathbf{F}_1 (imagem de referência) e \mathbf{F}_2 (imagem de entrada), com *pixels* representados por $f_1(x, y)$ e $f_2(x, y)$, respectivamente, em que x e y são as coordenadas desses *pixels* nas imagens. Considere-se ainda que \mathbf{F}_2 trata-se de uma versão transladada de \mathbf{F}_1 , conforme

$$f_2(x, y) = f_1(x - x_0, y - y_0), \quad (2.1)$$

em que x_0 e y_0 são respectivamente os deslocamentos relativos entre \mathbf{F}_1 e \mathbf{F}_2 no sentido das colunas e das linhas. Suas respectivas transformadas de Fourier, $\tilde{\mathbf{F}}_1$ e $\tilde{\mathbf{F}}_2$, podem ser relacionadas pela propriedade do deslocamento como (SMITH, 2002):

$$\tilde{f}_2(\xi, \eta) = e^{-j2\pi(\xi x_0 + \eta y_0)} \tilde{f}_1(\xi, \eta), \quad (2.2)$$

em que ξ e η são variáveis que denotam os componentes de frequência nas duas dimensões do domínio transformado. A fração oriunda de se isolar a exponencial em (2.2) resulta em um problema de deconvolução. Entretanto, considerando-se as imagens como sinais aleatórios, faz mais sentido que esse problema seja abordado de um ponto de vista estatístico. Assim sendo, as variáveis que se deseja estimar, x_0 e y_0 , podem ser determinadas a partir do espectro de potência cruzada (*cross-power spectrum* — ver Apêndice A) que consiste na Transformada de Fourier de Tempo Discreto (DTFT – *Discrete Time Fourier Transform*) da correlação-cruzada entre os sinais. Dada a similaridade entre a definição da correlação e da convolução, o espectro de potência cruzada normalizado pode ser definido direto de (2.2) (para mais detalhes, ver Apêndice A), conforme¹²:

$$\frac{\tilde{f}_2(\xi, \eta) \tilde{f}_1^*(\xi, \eta)}{|\tilde{f}_2(\xi, \eta) \tilde{f}_1^*(\xi, \eta)|} = e^{-j2\pi(\xi x_0 + \eta y_0)} \quad (2.3)$$

¹Cabe ressaltar que em (REDDY; CHATTERJI, 1996) e em (KUGLIN; HINES, 1975) o sinal do expoente da exponencial complexa em (2.2) e em (2.3) é inconsistente com a álgebra apresentada neste trabalho, que se baseia na substituição do lado direito de (2.2) no lado esquerdo de (2.3).

²Note-se que a normalização em (2.3) implica em se eliminar o módulo do produto $\tilde{f}_2(\xi, \eta) \tilde{f}_1^*(\xi, \eta)$ e se trabalhar apenas com a fase.

em que $\tilde{f}_1^*(\xi, \eta)$ é o complexo conjugado de $\tilde{f}_1(\xi, \eta)$. Sabendo-se que a transformada inversa de Fourier da exponencial complexa no lado direito de (2.3) é

$$\mathcal{F}^{-1}\{e^{-j2\pi(\xi x_0 + \eta y_0)}\} = \delta(x - x_0, y - y_0) \quad (2.4)$$

tem-se, substituindo (2.4) em (2.3), que

$$r(x, y) = \mathcal{F}^{-1}\left\{\frac{\tilde{f}_2(\xi, \eta)\tilde{f}_1^*(\xi, \eta)}{|\tilde{f}_2(\xi, \eta)\tilde{f}_1^*(\xi, \eta)|}\right\} = \delta(x - x_0, y - y_0) \quad (2.5)$$

de onde se pode obter as variáveis desejadas x_0 e y_0 . A estimação dos parâmetros utilizando somente a fase das imagens é o que distingue o Método de Correlação de Fase dos demais métodos de registro de imagens e estimação de movimento.

2.4.1 Estimação de Movimentos com Translação e Rotação

A estimação de movimentos de rotação também pode ser realizada explorando-se a propriedade do deslocamento da transformada da Fourier (REDDY; CHATTERJI, 1996). Considerando uma imagem F_1 como sendo, agora, uma versão transladada e rotacionada de uma imagem F_2 , segue direto da transformação de coordenadas (GONZALES; WOODS, 2001)

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \theta_0 & \sin \theta_0 \\ -\sin \theta_0 & \cos \theta_0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \quad (2.6)$$

que

$$f_2(x, y) = f_1(x \cos \theta_0 + y \sin \theta_0 - x_0, -x \sin \theta_0 + y \cos \theta_0 - y_0). \quad (2.7)$$

Aplicando-se a DTFT em (2.7) tem-se (REDDY; CHATTERJI, 1996):

$$\tilde{f}_2(\xi, \eta) = e^{-j2\pi(\xi x_0 + \eta y_0)} \tilde{f}_1(\xi \cos \theta_0 + \eta \sin \theta_0, -\xi \sin \theta_0 + \eta \cos \theta_0). \quad (2.8)$$

Como é possível observar em (2.2), o efeito de um deslocamento no domínio da imagem se restringe a uma alteração na fase, no domínio transformado. A partir disso, conclui-se que alterações no módulo, entre \tilde{F}_1 e \tilde{F}_2 , em (2.8) resultam exclusivamente da rotação no domínio na imagem (assumindo-se o caso ideal descrito pela equação)³.

A literatura (REDDY; CHATTERJI, 1996; KELLER; AVERBUCH; ISRAELI, 2005; PAN; QIN; CHEN, 2009; LI et al., 2012; TONG et al., 2019) busca, então, estimar os parâmetros de rotação a partir do módulo de (2.8):

$$|\tilde{f}_2(\xi, \eta)| = |\tilde{f}_1(\xi \cos \theta_0 + \eta \sin \theta_0, -\xi \sin \theta_0 + \eta \cos \theta_0)|. \quad (2.9)$$

³Note-se que a aplicação da DTFT serve, neste caso, para isolar as informações de translação e rotação (nos componentes de fase e módulo, respectivamente).

Tomando-se esses módulos como sendo os sinais de entrada de um novo problema de registro de imagens, recai-se em um problema de estimação de um movimento puramente rotacional. O problema rotacional pode ser transformado em um problema translacional fazendo-se uma transformação do sistema de referência dos sinais para coordenadas polares, conforme:

$$\tilde{m}_2(\rho, \theta) = \tilde{m}_1(\rho, \theta - \theta_0) \quad (2.10)$$

em que $\tilde{m}_1(\cdot, \cdot) = |\tilde{f}_1(\cdot, \cdot)|$, $\tilde{m}_2(\cdot, \cdot) = |\tilde{f}_2(\cdot, \cdot)|$, ρ e θ correspondem às coordenadas de módulo e ângulo, respectivamente, no sistema de coordenadas polares que mapeiam os sinais (“imagens”) de entrada do novo problema de registro. Esse problema, agora sendo translacional, pode ser resolvido através de (2.5).

2.4.2 Estimação de Movimentos de Reescala

Por questões didáticas e de adequação à aplicação de interesse, nesta seção são consideradas alterações de escala com a mesma proporção para as direções das linhas e das colunas da imagem. A extrapolação para escalas distintas em cada uma das direções é trivial e pode ser obtida em (REDDY; CHATTERJI, 1996). Assim sendo, se \mathbf{F}_1 for uma réplica em escala de \mathbf{F}_2 , com fator de escala a para as direções horizontal e vertical,

$$f_2(x, y) = f_1(ax, ay), \quad (2.11)$$

de acordo com a propriedade de escala da DTFT, as transformadas de \mathbf{F}_1 e \mathbf{F}_2 são relacionadas por (REDDY; CHATTERJI, 1996):

$$\tilde{f}_2(\xi, \eta) = \frac{1}{a^2} \tilde{f}_1(\xi/a, \eta/a). \quad (2.12)$$

Ao representar o sistema de referência (ξ, η) , em (2.12), em escala logarítmica, a mudança de escala entre $\tilde{\mathbf{F}}_1$ e $\tilde{\mathbf{F}}_2$ pode ser reduzida a um movimento de translação, conforme a relação de correspondência a seguir:

$$\tilde{f}_2(\ln \xi, \ln \eta) \hat{=} \tilde{f}_1(\ln \xi - \ln a, \ln \eta - \ln a). \quad (2.13)$$

Note-se que o fator de multiplicação $1/a^2$ pode ser ignorado na estimação da translação, visto que ela se baseia em correlação estatística. Nesse caso, uma diferença (especialmente invariante) na escala de amplitude entre dois sinais não altera a correlação entre eles.

O deslocamento $\ln a$ pode ser encontrado pela técnica de correlação de fase e a escala a pode ser calculada a partir do deslocamento $\ln a$, sendo

$$a = e^{\ln a} \quad (2.14)$$

em que e é a base do logaritmo natural.

2.4.3 Estimação de Movimentos com Translação, Rotação e Reescala

Considerando que \mathbf{F}_2 seja uma réplica transladada, rotacionada e redimensionada de \mathbf{F}_1 ,

$$f_2(x, y) = f_1 [a(x \cos \theta_0 + y \sin \theta_0 - x_0), a(-x \sin \theta_0 + y \cos \theta_0 - y_0)] \quad (2.15)$$

as transformadas de \mathbf{F}_2 e \mathbf{F}_1 são relacionadas por

$$\tilde{f}_2(\xi, \eta) = e^{-j2\pi(\xi x_0 + \eta y_0)} \frac{1}{a^2} \tilde{f}_1 \left(\frac{\xi \cos \theta_0 + \eta \sin \theta_0}{a}, \frac{-\xi \sin \theta_0 + \eta \cos \theta_0}{a} \right). \quad (2.16)$$

sendo seus módulos, então, relacionados por

$$\tilde{m}_2(\xi, \eta) = \frac{1}{a^2} \tilde{m}_1 \left(\frac{\xi \cos \theta_0 + \eta \sin \theta_0}{a}, \frac{-\xi \sin \theta_0 + \eta \cos \theta_0}{a} \right) \quad (2.17)$$

Assim sendo, a representação polar de $\tilde{\mathbf{M}}_2$ pode ser estimada a partir de

$$\rho_2 = \sqrt{\xi^2 + \eta^2} \quad (2.18)$$

e

$$\theta_2 = \tan^{-1} \left(\frac{\eta}{\xi} \right) \quad (2.19)$$

Já a representação polar de $\tilde{\mathbf{M}}_1$ pode ser estimada a partir de

$$\rho_1 = \sqrt{\left(\frac{\xi \cos \theta_0 + \eta \sin \theta_0}{a} \right)^2 + \left(\frac{-\xi \sin \theta_0 + \eta \cos \theta_0}{a} \right)^2} \quad (2.20)$$

e de

$$\theta_1 = \tan^{-1} \left(\frac{-\xi \sin \theta_0 + \eta \cos \theta_0 / a}{\xi \cos \theta_0 + \eta \sin \theta_0 / a} \right) = \tan^{-1} \left(\frac{-\xi \sin \theta_0 + \eta \cos \theta_0}{\xi \cos \theta_0 + \eta \sin \theta_0} \right) \quad (2.21)$$

Segue direto de (2.20) que

$$\rho_1 = \frac{1}{a} \sqrt{\xi^2 (\cos^2 \theta_0 + \sin^2 \theta_0) + \eta^2 (\cos^2 \theta_0 + \sin^2 \theta_0)} \quad (2.22)$$

e, portanto, a partir da identidade pitagórica $\cos^2 \theta_0 + \sin^2 \theta_0 = 1$, ρ_1 pode ser reescrito como

$$\rho_1 = \frac{1}{a} \sqrt{\xi^2 + \eta^2} = \frac{\rho_2}{a} \quad (2.23)$$

Dividindo-se o numerador e o denominador, em (2.21), por $\xi \cos \theta_0$, têm-se que

$$\theta_1 = \tan^{-1} \left(\frac{\frac{\eta}{\xi} - \frac{\sin \theta_0}{\cos \theta_0}}{1 + \frac{\eta \sin \theta_0}{\xi \cos \theta_0}} \right) \quad (2.24)$$

Sendo

$$\tan \theta_0 = \frac{\sin \theta_0}{\cos \theta_0}$$

(2.24) pode ser reescrita como

$$\theta_1 = \tan^{-1} \left(\frac{\frac{\eta}{\xi} - \tan \theta_0}{1 + \frac{\eta}{\xi} \tan \theta_0} \right) \quad (2.25)$$

Manipulando-se algebricamente (2.25), chega-se a

$$\theta_1 = \tan^{-1} \left\{ \frac{\tan \left[\tan^{-1} \left(\frac{\eta}{\xi} \right) \right] - \tan \theta_0}{1 + \tan \left[\tan^{-1} \left(\frac{\eta}{\xi} \right) \right] \tan \theta_0} \right\} \quad (2.26)$$

em que (ver (2.19))

$$\tan^{-1} \left(\frac{\eta}{\xi} \right) = \theta_2 \quad (2.27)$$

e, portanto,

$$\theta_1 = \tan^{-1} \left(\frac{\tan \theta_2 - \tan \theta_0}{1 + \tan \theta_2 \tan \theta_0} \right). \quad (2.28)$$

Considerando-se a identidade trigonométrica

$$\frac{\tan(a) - \tan(b)}{1 + \tan(a) \tan(b)} = \tan(a - b)$$

(2.28) pode ser simplificada para

$$\theta_1 = \tan^{-1} [\tan(\theta_2 - \theta_0)] = \theta_2 - \theta_0. \quad (2.29)$$

Então, em suma, $\rho_1 = \rho_2/a$ e $\theta_1 = \theta_2 - \theta_0$ e, em face disso, os espectros de magnitude de Fourier das imagens, na representação polar, podem ser relacionados por

$$\tilde{m}_2(\rho, \theta) = \tilde{m}_1(\rho/a, \theta - \theta_0) \quad (2.30)$$

e na representação log-polar são então relacionados por

$$\tilde{m}_2(\ln \rho, \theta) = \tilde{m}_1(\ln \rho - \ln a, \theta - \theta_0). \quad (2.31)$$

Com base em (2.31) e na técnica de correlação de fase, é possível obter a escala a e o ângulo θ_0 . Uma vez obtidos os valores de escala e ângulo, Reddy e Chatterji (1996) sugerem que a

imagem com a resolução mais alta seja redimensionada e rotacionada por esses valores para que a magnitude do movimento de translação seja estimada usando-se novamente a técnica de correlação de fase nas imagens resultantes. A redução de dimensão da maior imagem, em vez do aumento de dimensão da menor imagem, provavelmente é sugerido em razão de que imagens com maior resolução potencialmente contêm componentes de frequências mais altas. Ao se sub-amostrar a imagem maior, recai-se na comparação de duas imagens com a mesma limitação em frequência. Ao contrário, super-amostrando-se a imagem de menor resolução, recai-se na possibilidade de registro entre uma imagem com componentes de alta-frequência (imagem de alta resolução) com uma sem esses componentes (imagem de baixa resolução super-amostrada), o que pode reduzir o desempenho do algoritmo de registro.

2.5 QUESTÕES ACERCA DA IMPLEMENTAÇÃO DO MÉTODO

Nesta seção são discutidas algumas questões apontadas por diferentes autores que implementaram o Método de Correlação de Fase. Tópicos que merecem destaque são:

- Conversão de coordenadas cartesianas para polares — O espectro de magnitude log-polar de Fourier (2.31) é usado no lugar do espectro de magnitude polar de Fourier (2.30). Somente os dois quadrantes superiores do espectro de magnitude logarítmica são utilizados para o mapeamento para o plano polar. Isso ocorre pelo fato de o espectro de Fourier ser conjugado simétrico para sequências reais, ou seja (REDDY; CHATTERJI, 1996),

$$\tilde{f}(\xi, \eta) = \tilde{f}^*(-\xi, -\eta) \quad (2.32)$$

e, portanto,

$$|\tilde{f}(\xi, \eta)| = |\tilde{f}(-\xi, -\eta)|. \quad (2.33)$$

- Filtro passa-altas — Reddy e Chatterji (1996) propõem (sem razão explícita) que, antes de os espectros de magnitude logarítmica de Fourier serem mapeados para o plano polar logarítmico, sejam multiplicados por

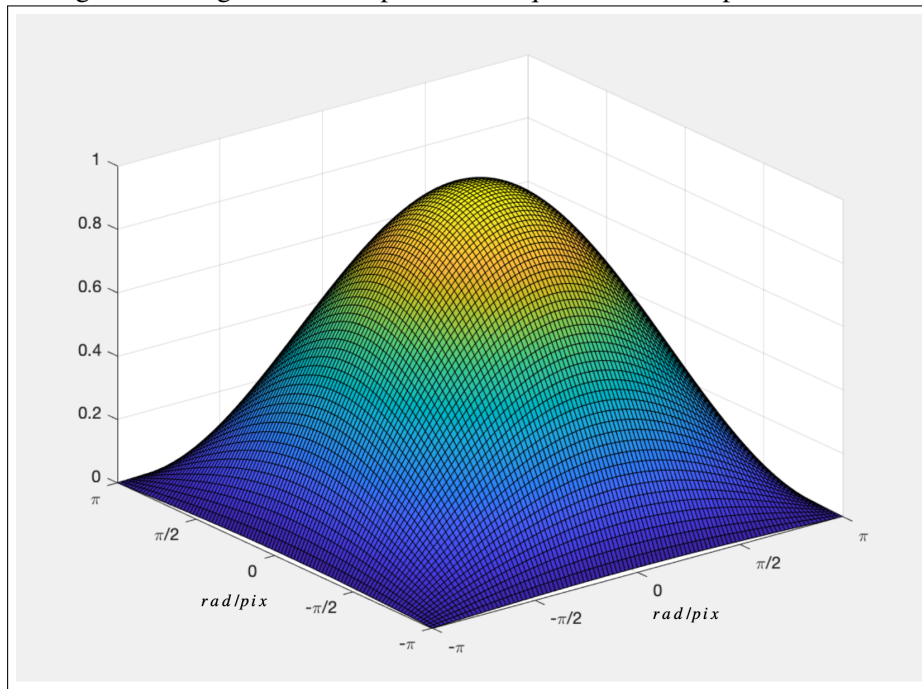
$$\tilde{h}(\xi, \eta) = [1.0 - \tilde{x}(\xi, \eta)][2.0 - \tilde{x}(\xi, \eta)] \quad (2.34)$$

em que

$$\tilde{x}(\xi, \eta) = \left[\cos\left(\frac{\xi}{2}\right) \cos\left(\frac{\eta}{2}\right) \right] \quad (2.35)$$

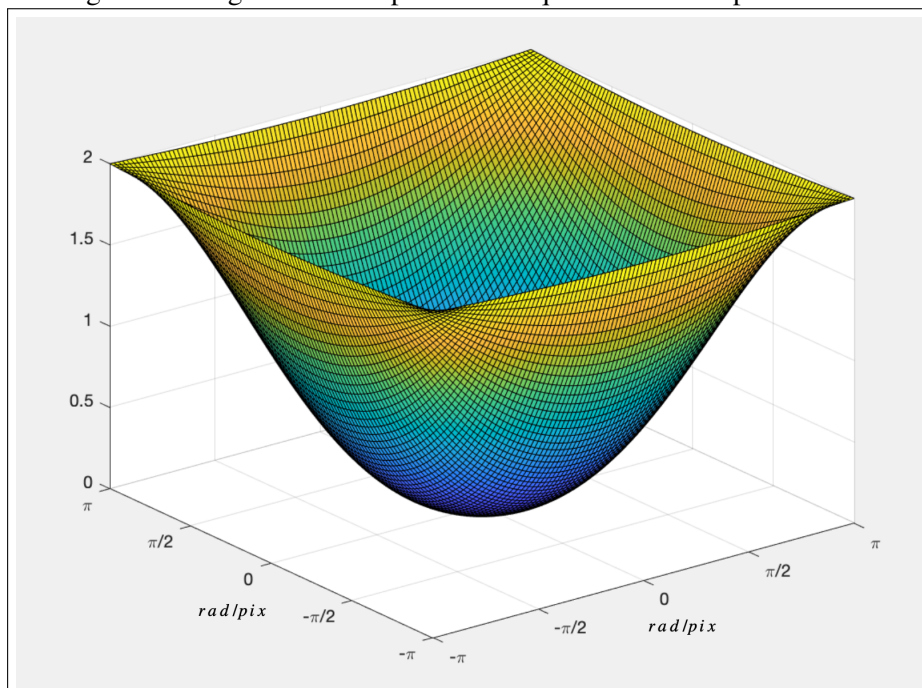
e $-\pi \leq \xi, \eta \leq \pi$. A magnitude da resposta em frequência do filtro \tilde{H} , caracterizado em (2.34), pode ser visualizada na Figura 3 e a magnitude de \tilde{X} , caracterizado em (2.35), pode ser visualizada na Figura 2. Em face da ausência de uma justificativa para a aplicação desse filtro passa-altas, esse assunto será abordado novamente no Capítulo 3.

Figura 2 – Magnitude da resposta em frequência do filtro passa-baixas



Fonte: Autor (2020).

Figura 3 – Magnitude da resposta em frequência do filtro passa-altas



Fonte: Adaptado de (LI et al., 2012).

- Transformação da imagem — O mapeamento do sistema de coordenadas da imagem do plano cartesiano para o plano log-polar geralmente resulta em coordenadas (índices) que não são inteiras. Nesses casos, Reddy e Chatterji (1996) sugerem que seja realizada uma interpolação linear entre os *pixels* vizinhos às coordenadas resultantes.
- Cálculo do espectro de Fourier — Reddy e Chatterji (1996) salientam que os espectros de Fourier das imagens podem ser calculados simultaneamente através de uma única FFT 2-D, aproveitando o fato de que cada matriz de dados de imagem possui apenas valores reais. Isso pode ser conseguido representando uma imagem como parte Real e a outra imagem como parte Imaginária de uma sequência complexa. Suas transformações podem ser separadas da transformação da sequência complexa por separação “par-ímpar”. Isto pode ser encontrado na literatura também como propriedade conjugada de Fourier ou simetria conjugada de Fourier.
- Reddy e Chatterji (1996) utilizaram imagens de tamanho 256×256 e a representação no plano log-polar também com tamanho de 256×256 . Note-se que nesse tipo de mapeamento a dimensão correspondente ao módulo (ρ) não se estende por mais do que a metade da maior dimensão da imagem (portanto, 128 *pixels*), assumindo-se a origem do sistema de referência no centro da imagem. A partir do aumento do Domínio de ρ (no caso, de 128 para 256), é possível considerar imagens de entrada com dimensões maiores que as da imagem de referência. As equações apresentadas pelos autores para a determinação das coordenadas de módulo e de ângulo são:

$$\rho = (1,044)^{1/2} \quad (2.36)$$

e

$$\theta = \frac{180y}{256} \quad (2.37)$$

em que a localização (coluna e linha) do pico da transformada inversa da fase do espectro de potência cruzada de Fourier são respectivamente dados por x e y . É possível perceber que, de acordo com (2.36), Reddy e Chatterji (1996) apresentam ρ como uma constante. Visto que a diferença de escala entre as imagens não é constante, constata-se que existe um equívoco nesta equação. Esse mapeamento é detalhado e esclarecido no Apêndice B.

- Considerando-se a propriedade de periodicidade dos sinais envolvidos⁴, um movimento relativo de translação horizontal, x_0 , entre duas imagens de tamanho $N \times N$ pode ser idealmente estimado como $\hat{x}_0 = x_0$ ou como $\hat{x}_0 = N - x_0$, dependendo da imagem considerada como imagem de entrada e imagem de referência. Se a imagem de entrada estiver deslocada para a direita em relação à imagem de referência ($x_0 > 0$) o pico de

⁴Sendo a implementação da Transformada de Fourier de Tempo Discreto feita pela Série de Fourier de Tempo Discreto, o sinal transformado pode ser interpretado como sendo um período de um sinal periódico (OPPENHEIM; WILLSKY; NAWAB, 1996).

correlação ocorrerá em um elemento x_0 (portanto positivo) da matriz formada por (2.5). Invertendo-se as imagens de entrada e de referência, agora com um deslocamento à esquerda ($x_0 < 0$), o pico de correlação seria equivalentemente encontrado no elemento $-x_0$ (negativo) da mesma matriz. Considerando-se a periodicidade dos sinais, esse pico será então encontrado em $\hat{x}_0 = N - x_0$. Cabe ressaltar que, na prática, os sinais não são periódicos. Sendo assim, qualquer estimativa \hat{x}_0 deve ser avaliada de forma a verificar se corresponde a x_0 para a direita ou a $N - x_0$ para a esquerda.

- De forma similar ao que ocorre na estimação de movimentos de translação, ao estimar o ângulo de rotação entre duas imagens, θ_0 , essa estimativa pode resultar em $\hat{\theta}_0 = \theta_0$ ou em $\hat{\theta}_0 = \theta_0 + 180^\circ$. O ângulo de 180° , neste caso, resulta da propriedade de simetria do módulo DTFT, definida em (2.33). No caso de sinais bidimensionais, então, o módulo do espectro (de onde se estima a rotação) é indistinguível de quando girado em 180° no entorno da origem. Assim como na translação, um ângulo negativo ($-\theta_0$) pode ser estimado a partir de $\hat{\theta}_0 = 360^\circ - \theta_0$. Porém, diferente do caso translacional, em que há diferença entre deslocar uma imagem (não periódica) em x_0 pixels para a direita ou em $N - x_0$ pixels para a esquerda, rotacionar o espectro em $-\theta_0$ é equivalente a rotacioná-lo em $360 - \theta_0$, não sendo necessário, portanto, o teste que avalia o sentido da rotação. Observe-se, por fim, que o ângulo de rotação é estimado no espaço log-polar, transformando-se em um problema de estimação de translação naquele plano.
- Influência do *Aliasing* — Stone, Tao e McGuire (2003) investigaram os fatores que diminuem a precisão do registro de imagens no domínio da frequência e destacam o *aliasing* dependente de rotação como um dos principais fatores. Como possível solução para este problema, apontam o corte das baixas frequências, pois afirmam que os efeitos do *aliasing* são mais fortes nessa banda, o que não faz sentido do ponto de vista do autor deste trabalho, como será discutido no Capítulo 3. Os autores propõem ainda o uso de um janelamento na imagem para eliminar altas frequências espúrias e apontam que Reddy e Chatterji (1996) utilizaram um filtro passa-altas para excluir frequências e ruídos indesejados.

Para reduzir a influência do *aliasing* no registro de imagens, Li et al. (2012) propuseram a *Multilayer-Pseudopolar Fractional Fourier Transform* (MPFFT). Essa transformada inclui uma transformação de sistema de coordenadas, e foi baseada em outras duas propostas similares, sendo elas a *Pseudo-Polar Fast Fourier Transform* (PPFFT), proposta por Keller, Averbuch e Israeli (2005), e a *Multilayer Fast Fourier Transform* (MLFFT), proposta por Pan, Qin e Chen (2009). Entretanto, é importante notar que, aparentemente, tanto Li et al. (2012) quanto Keller, Averbuch e Israeli (2005) e Pan, Qin e Chen (2009) utilizam apenas pares de imagens sintéticos em seus resultados. Quando isso não é evidente, como em Keller, Averbuch e Israeli (2005) e Pan, Qin e Chen (2009), uma avaliação mais cuidadosa das imagens leva a crer que o par de imagens a ser registrado consiste

em recortes (produzidos sinteticamente) de uma imagem natural. Sendo assim, o *aliasing* gerado pela rotação seguramente é um fator relevante em se tratando de pares de imagens que atendem às premissas definidas em (2.7). De outra forma, não necessariamente é um fator relevante quando uma das imagens contém inovações em relação à outra, ou seja, quando, por exemplo,

$$f_2(x, y) = f_1(x \cos \theta_0 + y \sin \theta_0 - x_0, -x \sin \theta_0 + y \cos \theta_0 - y_0) + s(x, y), \quad (2.38)$$

em que $s(x, y)$ é uma informação presente em \mathbf{F}_2 e não em \mathbf{F}_1 .

3 MÉTODO PROPOSTO

Neste trabalho é proposta a segmentação do objeto de interesse com base em registro de imagens em que, ao contrário do proposto em Cheraghi e Sheikh (2012), registra-se o objeto de interesse em vez do fundo da cena. O registro, por sua vez, fornece a posição e orientação do objeto presente na imagem de entrada em relação à imagem de referência. Para isso, propõe-se que a imagem de referência contenha somente o objeto, sobre um fundo preto. Ante a aplicação de interesse, essa segmentação é razoável pois só se faz necessária uma vez (durante o cadastro de um novo produto) e pode ser feita antes da aplicação em tempo-real, até mesmo de forma manual (em um editor de imagens).

O registro é proposto pelo método de Correlação de Fase, que estima no domínio da frequência o movimento relativo entre as imagens. Esse método se inicia pela estimação do ângulo de rotação, o que é feito com base no módulo dos espectros de Fourier do par de imagens. Para tanto, como discutido na Seção 2.4.1, os módulos dos espectros são mapeados para um plano (log-)polar¹ (ver Apêndice B). Esse mapeamento consiste em uma transformação de coordenadas entre elementos de duas matrizes de dados, ou seja, de uma matriz que corresponde aos dados (módulo do espectro de Fourier) no plano cartesiano, por ora chamada de \mathbf{C} , para outra que corresponde aos dados mapeados no plano polar, por ora chamada de \mathbf{P} . Assumindo-se que a matriz \mathbf{C} possui elementos indexados por $(\xi, \eta) \in \mathbb{Z}^2$ e a matriz \mathbf{P} possui elementos indexados por $(\rho, \theta) \in \mathbb{Z}^2$, segue que

$$(\rho_c, \theta_c) = T(\xi, \eta), \quad (3.1)$$

em que $T(\cdot)$ representa a função de transformação e (ρ_c, θ_c) representam o resultado da transformação de coordenadas. Note-se que, determinar (3.1) para toda coordenada (ξ, η) de \mathbf{C} tipicamente leva a valores $(\rho_c, \theta_c) \in \mathbb{R}^2$. A solução mais trivial para determinar (ρ, θ) a partir de (ρ_c, θ_c) seria fazer o arredondamento das coordenadas (ρ_c, θ_c) para o inteiro mais próximo, mas isso não garante que toda coordenada (ρ, θ) de \mathbf{P} (que esteja, de forma equivalente, dentro do domínio dos elementos de \mathbf{C}) receberá um valor oriundo da matriz \mathbf{C} (GONZALES; WOODS, 2001). Uma prática comum para lidar com isso é realizar o *mapeamento inverso* (GONZALES; WOODS, 2001)

$$(\xi_c, \eta_c) = T^{-1}(\rho, \theta), \quad (3.2)$$

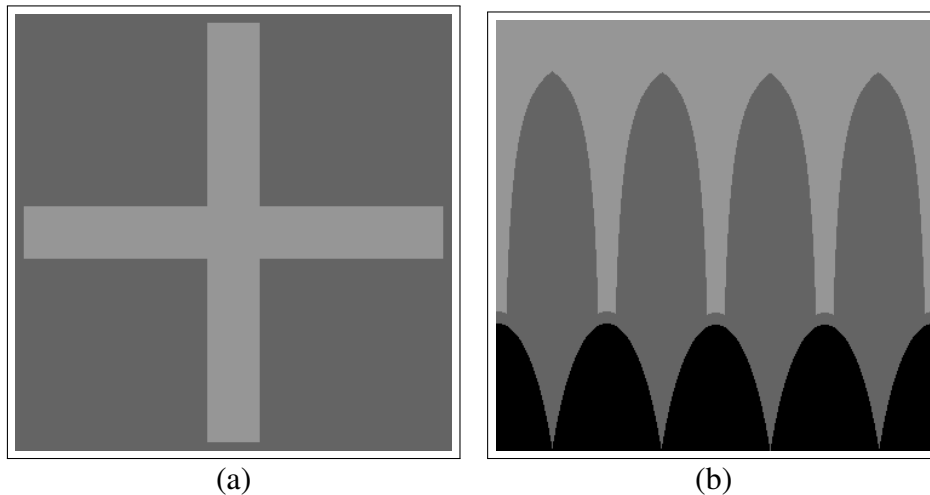
em que (ξ_c, η_c) representam o resultado da transformação inversa $T^{-1}(\cdot)$. Sendo $(\xi_c, \eta_c) \in \mathbb{R}^2$, para cada elemento (ρ, θ) de \mathbf{P} determina-se o endereço correspondente (ξ, η) de \mathbf{C} arredondando-se, da mesma forma que no mapeamento direto, as coordenadas (ξ_c, η_c) . Esse mapeamento garante, então, que a todo elemento (ρ, θ) seja atribuído um valor oriundo de \mathbf{C} . O arredondamento das coordenadas (ξ_c, η_c) pode ser interpretado como uma interpolação por vizinho

¹O mapeamento é feito para o plano polar, em vez de log-polar, quando não se considera variações de escala, como na aplicação de interesse deste trabalho.

mais próximo (KELLER; AVERBUCH; ISRAELI, 2005) e pode ser facilmente substituído por outras funções interpoladoras.

O mapeamento inverso é tradicionalmente adotado em algoritmos que contenham transformações geométricas (GONZALES; WOODS, 2001), a exemplo dos algoritmos de Correlação de Fase (KELLER; AVERBUCH; ISRAELI, 2005). A Figura 4 simboliza o módulo de um espectro (Figura 4(a)) transformado para o plano polar (Figura 4(b)) através de (3.2). As regiões que aparecem em preto na Figura 4(b) correspondem a regiões do plano polar que estão fora do domínio dos elementos do plano cartesiano, como ilustrado na Figura 14 do Apêndice B. Mesmo na presença de deslocamentos no plano polar, P , oriundos de rotações no plano cartesiano, C , os *pixels* pretos ilustrados na Figura 4(b), por serem estáticos, influenciam sobremaneira na correlação de fase, frequentemente resultando em uma estimativa de deslocamento nulo (rotação de 0° no plano cartesiano). A Figura 5 elucida esse problema. Na Figura 5(a) é apresentada uma versão rotacionada da Figura 4(a). A Figura 5(b) ilustra o mapeamento da Figura 5(a) para o plano polar. Embora os *pixels* cinza claro na Figura 5(b) sofram uma translação em relação aos da Figura 4(b), os *pixels* pretos apresentam-se nas mesmas posições em ambas as imagens. Esse problema pode ser resolvido desprezando-se as linhas inferiores das imagens nas figuras 4(b) e 5(b), o que equivale a mapear as imagens nas figuras 4(a) e 5(a) para o plano polar considerando-se apenas a região equivalente a de um círculo interno ao domínio dos seus elementos (*pixels*).

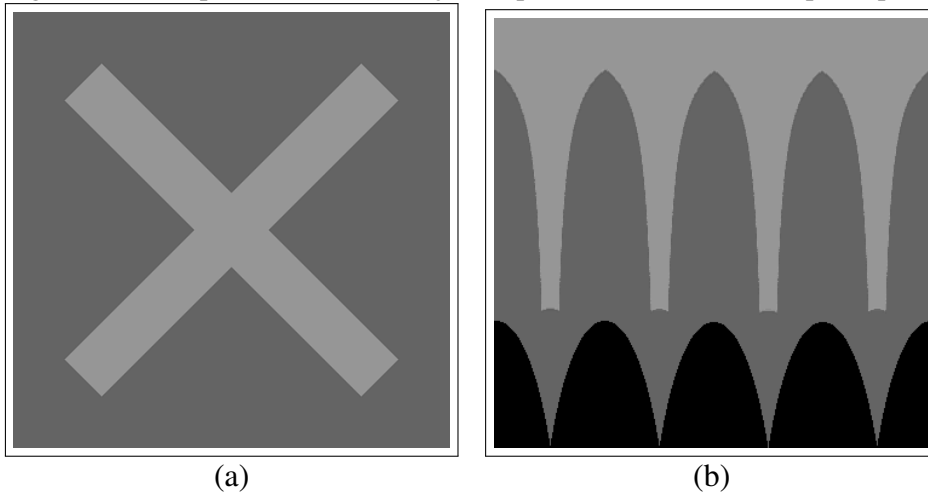
Figura 4 – Exemplo de transformação do plano cartesiano para o plano polar por meio de *mapeamento inverso*



(a) Ilustração do módulo de um espectro representado no plano cartesiano. (b) Módulo do espectro em (a) representado no plano polar. Fonte: Autor(2020).

Entretanto, mesmo mapeando-se apenas a região circular interna ao domínio dos elementos de C , a interpolação presente no mapeamento inverso estabelece um padrão de intensidades nos elementos de C que aparentemente tem passado despercebido pela comunidade científica nos últimos anos. Avaliando-se a transformação para o plano polar através do mapeamento direto, esse efeito é evidenciado. A Figura 6 ilustra o mapeamento do sinal apresentado na Figura 4(a)

Figura 5 – Exemplo de efeito da rotação no plano cartesiano sobre o plano polar



(a) Ilustração do módulo de um espectro representado no plano cartesiano. (b) Módulo do espectro em (a) representado no plano polar. Fonte: Autor(2020).

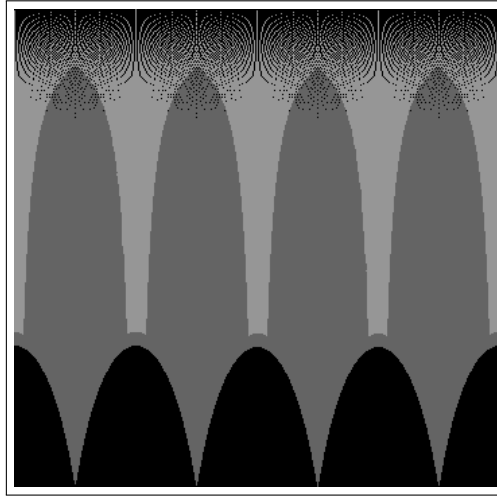
para o plano polar, através de (3.1). Nela, é possível notar uma malha de pontos pretos. Esses pontos, sem correspondente (ξ, η) atribuído pelo mapeamento direto, representam elementos para os quais as intensidades terão valores fortemente correlacionados com seus vizinhos (com correspondência), no mapeamento inverso. Note-se que as linhas superiores das matrizes de dados representadas nas figuras 4(b) e 6, com 360 elementos cada, considerando-se uma resolução de um grau, são mapeadas para a região central da matriz representada na Figura 4(b) (onde se localizam as baixas frequências do espectro de Fourier), em uma vizinhança de poucos ($\simeq \pi\rho^2$) elementos.

A região de baixas frequências do espectro de Fourier produzirá, portanto, uma forte correlação mesmo entre imagens de diferentes conteúdos. Essas correlações, estabelecidas pela característica mapeamento, é a razão de estimativas equivocadas de deslocamento nulo reportadas na literatura (KELLER; AVERBUCH; ISRAELI, 2005).

Uma solução para esse problema seria desprezar-se as primeiras linhas do plano (log-)polar, assim como desprezam-se as últimas. Esse “janelamento” do espectro explica o ganho de desempenho obtido pela filtragem passa-altas proposta de forma empírica em (REDDY; CHATTERJI, 1996). Entretanto, as baixas frequências podem carregar informações importantes a um registro acurado (imagens substancialmente homogêneas/alisadas podem depender das características de baixas frequências do espectro para um correto alinhamento), e a desconsideração dessa parte do espectro de Fourier pode levar o método de Correlação de Fase a uma perda substancial de robustez.

Para evitar tal perda de informação, neste trabalho é proposto que a rotação seja estimada no plano cartesiano (sem o mapeamento (log-)polar). Para tanto, o módulo do espectro de referência, $\tilde{M}_1 = |\tilde{F}_1|$, pode ser previamente rotacionado 360° em passos $\Delta\theta$, e, para cada ângulo de rotação, $\theta_n = n\Delta\theta$, em que $n \in \mathbb{Z}$, $0 \leq n \leq N - 1$ e $N = 360^\circ/\Delta\theta$, calcula-se o coeficiente de correlação, $\rho(\theta)$, de forma que o ângulo de rotação do módulo do espectro de

Figura 6 – Exemplo de transformação do plano cartesiano para o plano polar por meio de *mapeamento direto*



Fonte: Autor (2020).

entrada, $\tilde{\mathbf{M}}_2 = |\tilde{\mathbf{F}}_2|$, possa ser estimado conforme

$$\hat{\theta} = \arg \max_{\theta} \rho(\theta), \quad (3.3)$$

em que $\rho(\theta)$ pode ser estimado por (MANOLAKIS; INGLE; KOGON, 2005) como

$$\hat{\rho}(\theta) = \frac{\sum_{\xi} \sum_{\eta} [\tilde{m}_r(\xi, \eta, \theta) - \mu_{\tilde{m}_r}(\xi, \eta, \theta)] [\tilde{m}_2(\xi, \eta) - \mu_{\tilde{m}_2}(\xi, \eta)]}{\sqrt{\sum_{\xi} \sum_{\eta} [\tilde{m}_r(\xi, \eta, \theta) - \mu_{\tilde{m}_r}(\xi, \eta, \theta)]^2} \sqrt{\sum_{\xi} \sum_{\eta} [\tilde{m}_2(\xi, \eta) - \mu_{\tilde{m}_2}(\xi, \eta)]^2}}, \quad (3.4)$$

sendo

$$\tilde{m}_r(\xi, \eta, \theta) = \tilde{m}_1(\xi \cos \theta + \eta \sin \theta, -\xi \sin \theta + \eta \cos \theta) \quad (3.5)$$

os elementos da matriz $\tilde{\mathbf{M}}_1$ rotacionada. Note-se ainda que, para fins de implementação, as diversas versões rotacionadas, $\tilde{\mathbf{M}}_r(\theta)$, do módulo do espectro de referência, $\tilde{\mathbf{M}}_1$, assim como as estimativas de suas estatísticas de primeira e segunda ordem presentes em (3.5), podem ser estimadas *a priori* e armazenadas em memória.

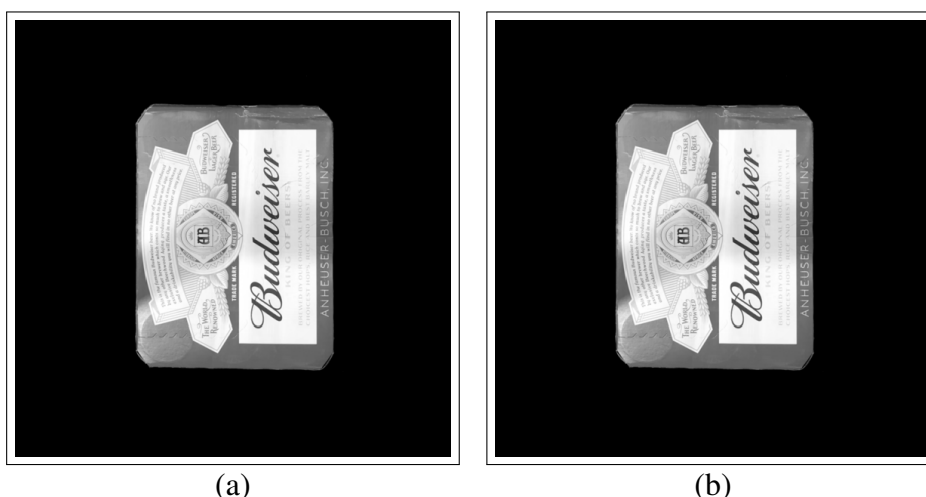
De forma a evitar o tratamento de fronteiras demandado pela rotação em (3.5), é proposto que a sequência de correlação seja estimada considerando-se apenas os elementos correspondentes ao maior círculo centrado no domínio dos espectros $\tilde{\mathbf{M}}_r(\theta)$ e $\tilde{\mathbf{M}}_2$. Para tanto, os espectros podem ser ponderados por uma janela circular com diâmetro correspondente à menor dimensão do espectro, sendo preservados os elementos no interior do círculo e anuladas as frequências externas ao círculo. Isso equivale a desprezar as últimas linhas do plano polar, como é habitualmente feito.

4 AVALIAÇÃO DO MÉTODO PROPOSTO

Neste capítulo são apresentados os testes realizados e resultados obtidos com o método proposto. Para que o erro de estimação de movimento possa ser avaliado, é necessário que se conheça o movimento verdadeiro, relativo, entre cada imagem de referência e de entrada. Por este motivo, as imagens utilizadas nas simulações a seguir são geradas sinteticamente. Inicialmente são utilizadas imagens que atendem idealmente aos pressupostos do método. Gradualmente, não-idealidades¹ são inseridas no processo, possibilitando uma avaliação de robustez, até que o método seja avaliado mediante condições de aplicação próximas às reais. Ao final, uma simulação é apresentada usando-se um par de imagens reais (naturais), de forma a ilustrar o desempenho do método sob condições reais. Ao todo são então apresentados 5 testes:

- I. *Objetos iguais sobre fundo preto* — O objetivo deste teste é avaliar o desempenho do registro mediante condições idealizadas. Nele, as imagens de entrada e de referência são iguais, a menos pela transformação geométrica sob avaliação (translação e rotação), e possuem fundo preto (o que, sob as condições simuladas, elimina a presença de inovações). Apenas mediante movimentos de rotação pode-se ter a presença de uma não idealidade referente ao *aliasing* dependente de rotação discutido no último item da Seção 2.5. Na Figura 7(a) é apresentada a imagem de referência (sempre a mesma para todos os testes). Na Figura 7(b) é apresentada a imagem de entrada que, embora seja exatamente a mesma que a apresentada em (a), durante os testes sofre translações e rotações.

Figura 7 – Exemplos de imagens de referência e de entrada usadas no teste I (objetos iguais sobre fundo preto).



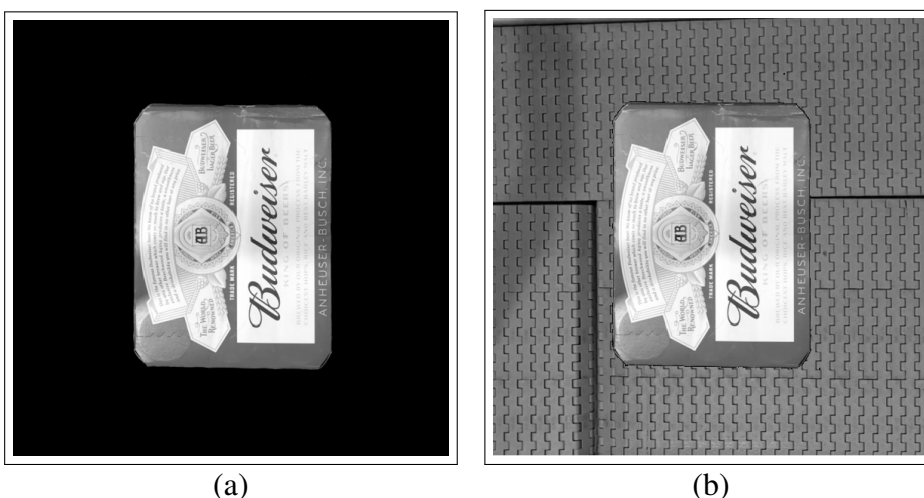
(a) Imagem de referência. (b) Imagem de entrada. Fonte: Autor(2020).

- II. *Objetos iguais sobre fundos distintos* — O objetivo deste teste é avaliar o desempenho do registro na presença de inovações. O objeto de interesse, tanto na imagens de en-

¹Trata-se, a partir deste ponto, por não-idealidade toda e qualquer característica do par de imagens a ser registrado que, considerada na simulação, não atenda à premissa em (2.15).

trada quanto na de referência, é idêntico (*pixels* com intensidades idênticas), a menos por eventual interpolação necessária à implementação de movimentos de rotação. O fundo da imagem de entrada, entretanto, é alterado sinteticamente. Neste caso, enquanto a imagem de referência é mantida com fundo preto, conforme ilustrado na Figura 8(a), a imagem de entrada contém o objeto de interesse sobreposto (sinteticamente) a um fundo que retrata uma esteira transportadora, conforme ilustrado na Figura 8(b). O objeto da imagem de entrada sofre translações e/ou rotações (sintéticas) enquanto o fundo permanece estático.

Figura 8 – Exemplos de imagens de referência e de entrada usadas no teste II (objetos iguais sobre fundos distintos).



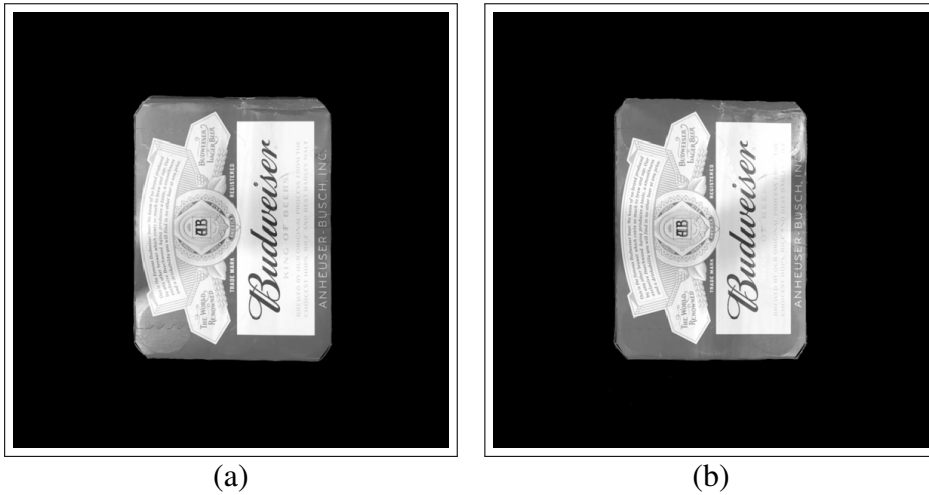
(a) Imagem de referência. (b) Imagem de entrada. Fonte: Autor (2020).

III. *Objetos distintos sobre fundo preto* — O objetivo deste teste é avaliar o desempenho do registro mediante um objeto de interesse com componente aleatório. Nele, as imagens de entrada e de referência possuem fundo preto, mas os objetos de interesse são distintos (imagens distintas, de dois fardos de um mesmo produto, como acontece em uma aplicação real). A imagem de referência e uma imagem de entrada são ilustradas, respectivamente, na Figura 9(a) e (b). É possível notar um aparente desgaste no canto superior esquerdo da embalagem na imagem de referência, em relação à imagem de entrada, o que retrata a diferença entre as aquisições. A recíproca é verdadeira em relação ao canto inferior direito. Além disso, os padrões de iluminação no centro da embalagem também são distintos.

IV. *Objetos distintos sobre fundos distintos* — O objetivo deste teste é avaliar o desempenho do registro na presença de inovações e de um objeto de interesse com componente aleatório. As imagens de entrada e de referência possuem objetos de interesse distintos, sendo na imagem de referência o objeto sobreposto a um fundo preto e na imagem de entrada o objeto sobreposto a um fundo que retrata uma esteira de transporte, conforme ilustrado na Figura 10.

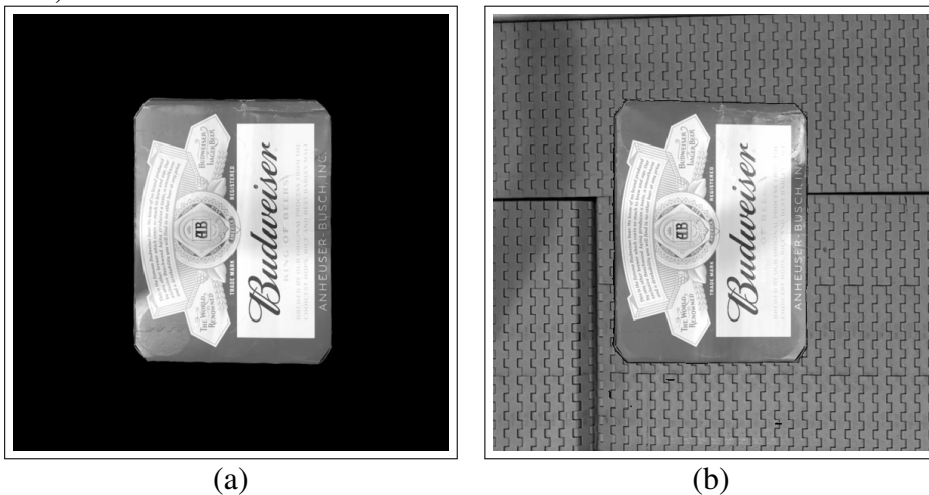
V. *Imagens naturais* — Teste realizado em condições reais de aplicação. A imagem de

Figura 9 – Exemplos de imagens de referência e de entrada usadas no teste III (objetos distintos sobre fundo preto).



(a) Imagem de referência. (b) Imagem de entrada. Fonte: Autor(2020).

Figura 10 – Exemplos de imagens de referência e de entrada usadas no teste IV (objetos distintos sobre fundos distintos).

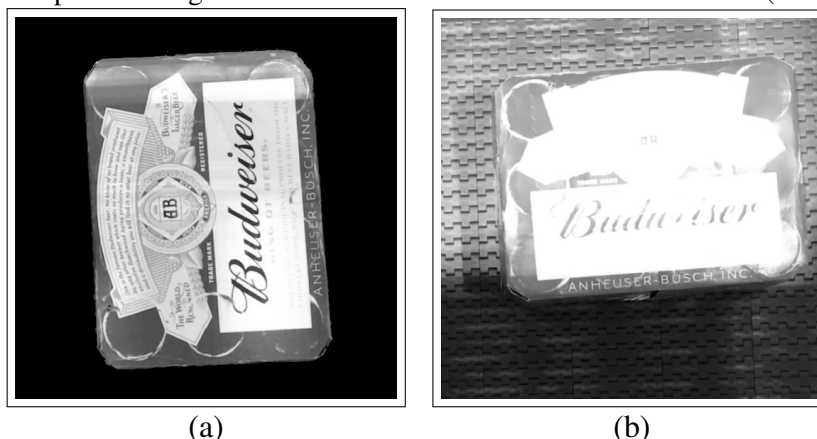


(a) Imagem de referência. (b) Imagem de entrada. Fonte: Autor(2020).

referência é adquirida *a priori* e o fundo removido manualmente. A imagem de entrada é adquirida em uma linha de envase, em que o objeto de interesse é disposto em posição aleatória sobre uma esteira transportadora, conforme ilustrado na Figura 11. Como é possível observar, de forma a ilustrar a robustez do método proposto, a imagem de entrada sofre forte interferências da iluminação.

Para uma melhor apresentação das contribuições deste trabalho, os resultados dos cinco conjuntos de testes são divididos em três grupos de simulações, nas três seções que se seguem. Na Seção 4.1 são apresentados os resultados dos testes I a IV considerando-se apenas movimentos de translação. Nesse caso, não impactam as alterações propostas neste trabalho sobre o método de Correlação de Fase. Trata-se apenas de uma avaliação da adequação do método, na forma em que se propõe o uso dele, à aplicação de interesse. Na Seção 4.2 são apresentados os resultados

Figura 11 – Exemplos de imagens de referência e de entrada usadas no teste V (imagens naturais).



(a) Imagem de referência. (b) Imagem de entrada. Fonte: Autor(2020).

dos testes I a IV considerando-se apenas movimentos de rotação. Nesse caso, impactam as alterações propostas neste trabalho sobre o método de Correlação de Fase. Por fim, na Seção 4.3 é apresentado um resultado de simulação referente ao teste V, considerando-se conjuntamente, neste caso, movimentos de rotação e de translação. Cabe ressaltar que em imagens naturais não é possível isolar com precisão apenas um tipo de movimento.

Em todas as simulações, as imagens sintéticas foram geradas respeitando-se amplitudes de translação e de rotação que mantêm o objeto de interesse integralmente dentro do domínio da imagem. As translações foram implementadas com deslocamentos sucessivos de um *pixel*, varrendo-se uma linha após a outra, e as rotações foram implementadas com passos de um grau, utilizando-se interpolação bicúbica quando necessário. Nos testes II e IV, apenas o objeto de interesse foi movimentado, enquanto o fundo foi mantido estático. Em todas as simulações, são utilizadas imagens de tamanho 575×575 . O movimento de rotação do objeto de interesse, nos testes I a IV, foi sintetizado com uma resolução de um grau, o que gerou 360 imagens de entrada (para cada teste) usadas na avaliação da estimação de rotação. O movimento de translação do objeto de interesse, nos testes I a IV, foi sintetizado por deslocamentos de passo unitário na direção das linhas e das colunas da imagem de entrada. Restringiu-se a translação de forma a manter o objeto de interesse dentro dos limites da imagem. Com isso, foram geradas 71.416 imagens para cada teste.

Ainda que os testes propostos sejam desempenhados com um único objeto de interesse, não é esperado que, por exemplo, uma alteração de formato ou na embalagem desse objeto venha a impactar no desempenho do algoritmo de registro. Os problemas de desempenho se dão, tipicamente, por inconsistências entre os sinais e suas modelagens. O que é tipicamente mais desvantajoso para o registro de imagens é a ausência de detalhes (altas frequências) nas imagens. Por exemplo, mediante embalagens brancas, é esperada uma redução no desempenho do registro. Ainda assim, em termos de comparação entre o método original e o método proposto, não se espera alteração de comportamento frente a diferentes conteúdos. Simulações exaustivas são sugeridas como trabalhos futuros.

4.1 ESTIMAÇÃO DE MOVIMENTOS DE TRANSLAÇÃO

Na Tabela 1 são apresentados os erros de estimação de translação obtidos nos teste I a IV, considerando-se 71.416 estimações para cada teste. É possível observar que a alteração do fundo da cena, que representa uma maior presença de inovações na imagem de entrada, gera maior influência sobre o desempenho do método de Correlação de Fase que a presença de variações de iluminação e de variações no aspecto físico do objeto de interesse. Note-se que, ainda que essas últimas variações consideradas sejam sutis, foram consideradas de acordo com a realidade de uma linha de produção, em que as embalagens estão tipicamente intactas e a iluminação pode ser, de certa forma, controlada.

Sob as condições do teste IV, os resultados obtidos representam um erro absoluto médio menor que 1% e um erro absoluto máximo de aproximadamente 2% em relação às dimensões das imagens. Ainda que esse nível de precisão possa parecer aceitável, cabe ressaltar que o algoritmo estima primeiro os movimentos de rotação. Em posse desse valor, registra as imagens em função da rotação encontrada para depois avaliar a translação. Assim, erros na estimação rotação, avaliados na próxima seção, impactam diretamente na correlação para movimentos translacionais.

Tabela 1 – Erros de estimação de posição, em *pixels*, mediante movimentos translacionais.

Teste	Erro absoluto médio	Erro absoluto máximo	Desvio Padrão
I – <i>Objetos iguais sobre fundo preto</i>	0,0000	0,0000	0,0000
II – <i>Objetos iguais sobre fundos distintos</i>	1,4835	4,2426	0,7533
III – <i>Objetos distintos sobre fundo preto</i>	0,0190	4,0000	0,2509
IV – <i>Objetos distintos sobre fundos distintos</i>	5,2564	12,0416	1,3127

Fonte: Autor (2022).

4.2 ESTIMAÇÃO DE MOVIMENTOS DE ROTAÇÃO

Nas Tabela 2 são apresentados os erros de estimação de rotação para o método proposto em (REDDY; CHATTERJI, 1996) (aqui tratado como *método original*) e para o método proposto, considerando-se os testes de I a IV, considerando-se 360 estimações para cada teste. Diferente do caso de movimentos translacionais, é possível observar que para movimentos rotacionais a alteração do fundo da cena, que representa uma maior presença de inovações na imagem de entrada, gera menor influência sobre o desempenho do método original de Correlação de Fase que a presença de variações de iluminação e de variações no aspecto físico do objeto de interesse. Sob as condições do teste IV, os resultados obtidos com o método original representam um erro

absoluto médio de aproximadamente 8% (30°) e um erro absoluto máximo de aproximadamente 43% (155°). Quando comparados os resultados do teste IV, o método proposto reduz o erro absoluto médio para aproximadamente 0,007% e o erro absoluto máximo para menos de 0,3%. Isso resulta em uma melhora de desempenho de cerca de três ordens de grandeza (1200 vezes) em termos de erro absoluto médio e de duas ordens de grandeza (155 vezes) em termos de erro absoluto máximo.

Tabela 2 – Erros de estimação de rotação (em graus)

Método	Teste	Erro absoluto médio	Erro absoluto máximo	Desvio Padrão
Original	I – <i>Objetos iguais sobre fundos pretos</i>	0,1564	0,3171	0,0910
	II – <i>Objetos iguais sobre fundos distintos</i>	0,1564	0,3206	0,0910
	III – <i>Objetos distintos sobre fundo preto</i>	0,2244	0,6307	0,1584
	IV – <i>Objetos Distintos sobre fundos distintos</i>	29,9905	155,3728	44,4078
Proposto	I – <i>Objetos iguais sobre fundos pretos</i>	0,0000	0,0000	0,0000
	II – <i>Objetos iguais sobre fundos distintos</i>	0,0000	0,0000	0,0000
	III – <i>Objetos distintos sobre fundo preto</i>	0,0000	0,0000	0,0000
	IV – <i>Objetos Distintos sobre fundos distintos</i>	0,0250	1,0000	0,1583

Fonte: Autor (2022).

4.3 RESULTADOS PARA MOVIMENTOS DE ROTAÇÃO E TRANSLAÇÃO

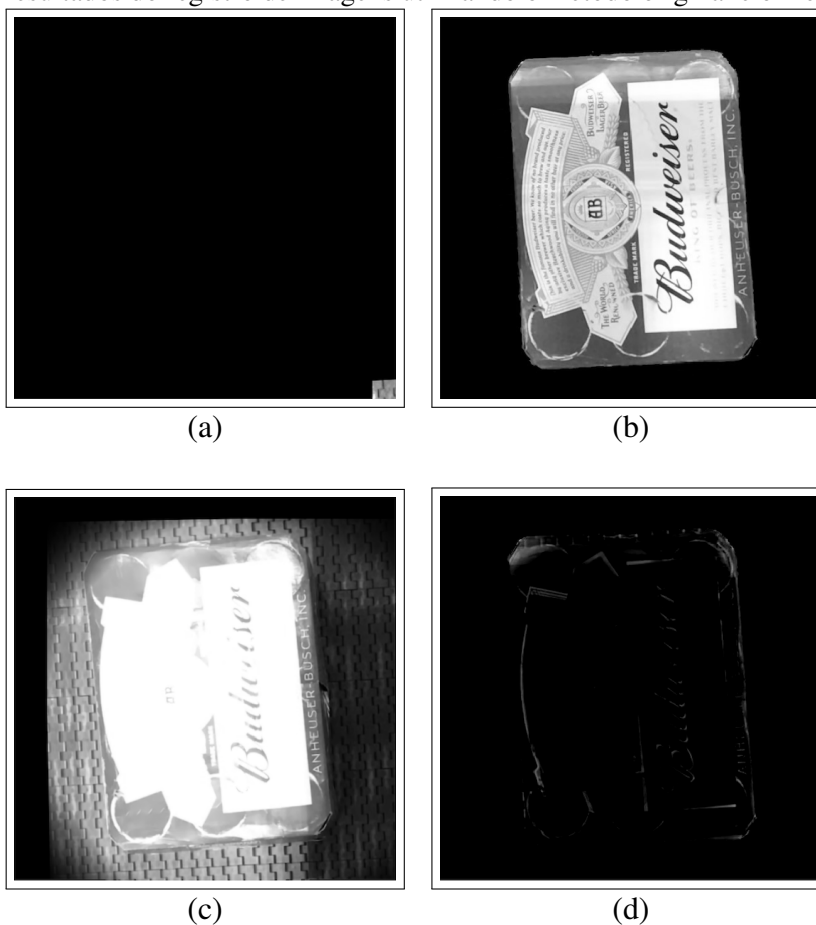
Conforme discutido, esta seção ilustra resultados referentes ao teste V, realizado com imagens naturais. Nessas condições não é possível determinar o erro de estimação de movimento em *pixels* e em graus, como nas seções anteriores, mas é possível realizar uma avaliação qualitativa desses erros.

O resultado do registro da imagem de entrada, ilustrada na Figura 11(b), com a imagem de referência (Figura 11(a)), com base no deslocamento estimado pelo método original é mostrado na Figura 12(a). Nela, o objeto de interesse foi registrado para fora do domínio da imagem (ver pequena parte do objeto no canto inferior direito da cena), isto é, o erro na estimação da translação foi na ordem de grandeza do próprio tamanho da imagem. O erro de registro é ilustrado na Figura 12(b). Note-se que esse erro, geralmente avaliado como o absoluto da diferença entre a imagem de referência e a imagem registrada, nesta simulação é avaliado apenas pelos valores positivos da diferença entre as duas imagens. Uma vez que a imagem de referência possui fundo

preto (níveis de intensidade iguais a zero), isso faz com que não sejam atribuídos valores de erro a regiões que não correspondam ao objeto de interesse na imagem de referência, facilitando a visualização. O resultado desse mesmo registro, quando realizado com o método proposto, é apresentado na Figura 12(c) e o respectivo erro de registro é mostrado na Figura 12(d).

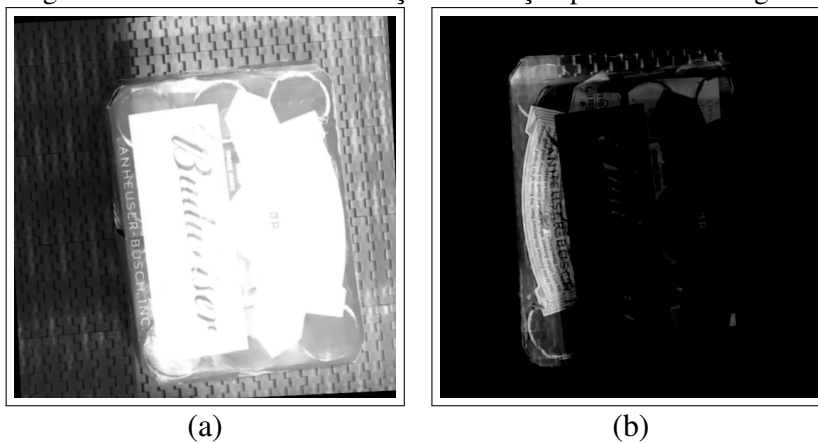
O resultado da etapa de estimação do movimento de rotação feita pelo método original é ilustrado pelo registro apresentado na Figura 13. É possível notar um erro de estimação de aproximadamente 180° em relação à imagem de referência. Esse erro na estimação da rotação leva, provavelmente, ao erro na estimação da translação apresentado na Figura 12(a).

Figura 12 – Resultados do registro de imagens utilizando o método original e o método proposto



(a) Objeto registrado pelo método original. (b) Erro de registro resultante do método original. (c) Objeto registrado pelo método proposto. (d) Erro de registro resultante do método proposto. Fonte: Autor(2020).

Figura 13 – Resultado da estimação de rotação pelo método original



(a) Imagem registrada. (b) Erro de registro. Fonte: Autor (2020).

5 CONCLUSÃO

Este trabalho foi fomentado pela Máquinas Sanmartin LTDA, uma das principais empresas do mundo na produção de máquinas dedicadas ao transporte, empacotamento e paletização e despaletização de bebidas. No ano de 2019 a Sanmartin inaugurou um Centro de Desenvolvimento e Inovação, o CDI Sanmartin, na cidade de Caxias do Sul, sendo ele um espaço com ampla área e infraestrutura para P,D&I, onde a maior parte desta pesquisa foi desenvolvida.

A detecção de objetos tem sido amplamente estudada na área de visão computacional, porém os métodos encontrados na literatura são capazes de determinar apenas a posição do objeto em uma imagem, e não a orientação (ângulo de rotação) em que esse objeto se apresenta. Além disso, os métodos disponíveis não costumam atingir níveis de acurácia demandados por aplicações industriais de envase e paletização.

Em face disso, neste trabalho foi proposto um método de estimação da posição e da orientação de objetos para aplicação em manipulação robótica em linhas de produção. Esse método se constitui como uma ferramenta capaz de lidar com uma ampla gama de objetos, potencialmente sem a necessidade de enclausuramento, e sem a necessidade de construção de conjuntos de treinamento. Para tanto, a solução proposta se baseia no registro de imagens por meio do método de Correlação de Fase.

Revisitando-se os objetivos deste trabalho, a proposta ora apresentada passa pela: (i) identificação da abordagem de registro como forma de estimar a posição e orientação do objeto de interesse na cena; (ii) pela seleção do método de Correlação de Fase como base para a solução; (iii) pelo desenvolvimento da solução para o problema em questão, que inclui a segmentação de imagens pelo registro do objeto de interesse e proposta de mudanças no método de Correlação de Fase; e (iv) conclui-se com a avaliação da solução proposta, que aponta melhoras no desempenho do método de registro em mais de 1000%.

Em se tratando do método de Correlação de Fase, amplamente discutido na literatura nas últimas décadas, o trabalho apresentado leva à seguinte **conclusão**: o mapeamento de coordenadas cartesianas para coordenadas polares usado na estimação de movimentos de rotação gera padrões de correlação entre as imagens que resultam na ocorrência de erros de registro severos e recorrentes. Para solucionar este problema é proposto que a estimação de movimentos rotacionais seja realizada no plano cartesiano. Os resultados apresentados mostram que esta proposta pode melhorar o desempenho do registro entre duas a três ordens de grandeza.

Por fim, como trabalhos futuros, sugere-se:

- Avaliar o impacto da característica dos sinais de entrada, a exemplo de alterações nas embalagens, no desempenho do registro;
- Implementar o método proposto em linha de produção, de forma que se possa obter uma avaliação quantitativa de desempenho, considerando-se os requisitos específicos da aplicação;

- Caracterizar o grau de inovações entre o par de imagens suportado pelo algoritmo de registro.

REFERÊNCIAS

- ANTIPOV, V. et al. Computer vision system for recognition and detection of color patterns in real-time task of robot control. **2017 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SINKHROINFO)**, p. 1–5, jul. 2017. <<https://ieeexplore.ieee.org/document/7997496>>, Acesso em: 01 jun. 2020.
- BAO, J.; YE, M. Head pose estimation based on robust convolutional neural network. *Cybernetics and Information Technologies*, v. 16, 12 2016.
- BROWN, L. G. A survey of image registration techniques. **ACM Computing Surveys**, Nova York, v. 24, n. 4, p. 325–376, dez. 1992. <<http://www.sci.utah.edu/~gerig/CS6640-F2010/p325-brown.pdf>>, Acesso em: 24 jan. 2020.
- BRUCE, J.; VELOSO, M. Fast and accurate vision-based pattern detection and identification. **2003 IEEE International Conference on Robotics and Automation**, p. 1277–1282, set. 2003. <<https://ieeexplore.ieee.org/document/1241768>>, Acesso em: 01 jun. 2020.
- CASTRO, E. D.; MORANDI, C. Registration of translated and rotated images using finite fourier transforms. **IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE**, PAMI-9, n. 5, p. 700–703, set. 1987. <<http://adessowiki.fee.unicamp.br/media/Attachments/courseIA369O1S2011/Ex13/phasecorrelation6.pdf>>, Acesso em: 25 jul. 2020.
- CHERAGHI, S. A.; SHEIKH, U. U. Moving object detection using image registration for a moving camera platform. **2012 IEEE International Conference on Control System, Computing and Engineering**, p. 355–359, nov. 2012. <<https://ieeexplore.ieee.org/document/6487170/>>, Acesso em: 7 ago. 2020.
- FEFILATYEV, S.; GOLDGOF, D.; LEMBKE, C. Tracking ships from fast moving camera through image registration. **2010 20th International Conference on Pattern Recognition**, p. 3500–3503, 2010.
- FENG, J.; YI-CAN, L. A fast and robust image registration algorithm. **2019 6th International Conference on Information Science and Control Engineering (ICISCE)**, p. 36–38, dez. 2019. <<https://ieeexplore.ieee.org/document/9107828>>, Acesso em: 27 jul. 2020.
- FONSECA, L.; KENNEY, C. Control point assessment for image registration. **XII Brazilian Symposium on Computer Graphics and Image Processing (Cat. No.PR00481)**, p. 1–8, out. 1999. <<https://ieeexplore.ieee.org/document/805717>>, Acesso em: 27 jul. 2020.
- FUKUNAGA, K.; FLICK, T. E. Classification error for a very large number of classes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, PAMI-6, n. 6, p. 779–788, 1984.
- GONZALES, R. C.; WOODS, R. E. **Digital Image Processing**. [S.l.]: Prentice Hall, 2001. ISBN 0-201-18075-8.
- HENEGHAN, C. et al. Retinal image registration using control points. **Proceedings IEEE International Symposium on Biomedical Imaging**, p. 349–352, jul. 2002. <<https://ieeexplore.ieee.org/document/1029265>>, Acesso em: 27 jul. 2020.

- JAHANANDISH, M. H.; FEY, N. P.; HOYT, K. Lower limb motion estimation using ultrasound imaging: A framework for assistive device control. **IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS**, v. 23, n. 6, p. 2505–2514, nov. 2019. <<https://ieeexplore.ieee.org/document/8606176>>, Acesso em: 01 jun. 2020.
- JIANG, X. et al. Robust feature matching for remote sensing image registration via linear adaptive filtering. *IEEE Transactions on Geoscience and Remote Sensing*, v. 59, n. 2, p. 1577–1591, 2021.
- JOSHI, A. J.; PORIKLI, F.; PAPANIKOLOPOULOS, N. Multi-class active learning for image classification. **2009 IEEE Conference on Computer Vision and Pattern Recognition**, p. 2372–2379, 2009.
- KELLER, Y.; AVERBUCH, A.; ISRAELI, M. Pseudopolar-based estimation of large translations, rotations, and scalings in images. **IEEE TRANSACTIONS ON IMAGE PROCESSING**, v. 14, n. 1, p. 12–22, jan. 2005. <<https://ieeexplore-ieee-org.ez314.periodicos.capes.gov.br/stamp/stamp.jsp?tp=&arnumber=1369326>>, Acesso em: 30 abr. 2020.
- KHARE, S. Finger gesture and pattern recognition based device security system. **2015 International Conference on Signal Processing and Communication (ICSC)**, p. 443–447, jul. 2015. <<https://ieeexplore.ieee.org/document/7150694>>, Acesso em: 24 jul. 2020.
- KIM, J. U.; RO, Y. M. Attentive layer separation for object classification and object localization in object detection. **2019 IEEE International Conference on Image Processing (ICIP)**, p. 3995–3999, set. 2019. <<https://ieeexplore.ieee.org/document/8803439>>, Acesso em: 7 ago. 2020.
- KUGLIN, C. D.; HINES, D. C. The phase correlation image alignment method. **Proceeding of IEEE International Conference on Cybernetics and Society**, Nova York, n. 4, p. 163–165, set. 1975. <<http://boutigny.free.fr/Astronomie/AstroSources/Kuglin-Hines.pdf>>, Acesso em: 24 jan. 2020.
- LI, Z. et al. Multilayer-pseudopolar fractional fourier transform approach for image registration. **2012 Eighth International Conference on Computational Intelligence and Security**, p. 323–327, nov. 2012. <<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6405854>>, Acesso em: 29 jan. 2020.
- MAI, T. et al. Large scale multi-class classification using latent classifiers. **2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)**, p. 1–6, 2015.
- MAINTZ, J. B. A.; VIERGEVER, M. A. A survey of medical image registration. **Medical Image Analysis**, v. 2, n. 1, p. 1–36, mar. 1998. <<https://www.sciencedirect.com/science/article/pii/S1361841501800268>>, Acesso em: 19 fev. 2020.
- MANOLAKIS, D.; INGLE, V.; KOGON, S. M. Statistical and adaptive signal processing. [S.l.]: Artech House, 2005. ISBN 1580536107.
- MENG, Z.; HAN, X.; LI, X. An automatic registration method for sar and optical images based on line extraction and control points selection. **2017 International Applied Computational Electromagnetics Society Symposium (ACES)**, p. 1–2, ago. 2017. <<https://ieeexplore.ieee.org/document/8051755>>, Acesso em: 27 jul. 2020.

NAG, S. Image registration techniques: A survey. **Technical Report, Department of Electrical Engineering, Jadavpur University**, p. 1–9, nov. 2017. <https://www.researchgate.net/publication/321342677_Image_Registration_Techniques_A_Survey>, Acesso em: 02 jun. 2020.

NGUYEN, K. et al. Bayesian multi-hyperplane machine for pattern recognition. **2018 24th International Conference on Pattern Recognition (ICPR)**, p. 609–614, nov. 2018. <<https://ieeexplore.ieee.org/document/8545139>>, Acesso em: 24 jul. 2020.

NILSBACK, M.; ZISSERMAN, A. Automated flower classification over a large number of classes. **2008 Sixth Indian Conference on Computer Vision, Graphics Image Processing**, p. 722–729, 2008.

OMAROV, B.; CHO, Y. I. Machine learning based pattern recognition and classification framework development. **2017 17th International Conference on Control, Automation and Systems (ICCAS 2017)**, p. 1–5, out. 2017. <<https://ieeexplore.ieee.org/document/8204282>>, Acesso em: 24 jul. 2020.

OPPENHEIM, A. V.; WILLSKY, A. S.; NAWAB, H. S. **Signals & Systems**. [S.l.]: Prentice-Hall, 1996. <https://web.itu.edu.tr/hulyayalcin/Signal_Processing_Books/Oppenheim_Signals_and_Systems.pdf>, Acesso em: 08 jun. 2020. ISBN 0138147574.

PAN, W.; QIN, K.; CHEN, Y. An adaptable-multilayer fractional fourier transform approach for image registration. **IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE**, v. 31, n. 3, p. 400–413, mar. 2009. <<https://ieeexplore-ieee-org.ez314.periodicos.capes.gov.br/stamp/stamp.jsp?tp=&arnumber=4483796>>, Acesso em: 30 abr. 2020.

PEI, W.; JIA, Y. Feature extraction and image registration in the final approaching phase of rendezvous and docking. **2018 37th Chinese Control Conference (CCC)**, p. 9913–9917, jul. 2018. <<https://ieeexplore.ieee.org/document/8483006>>, Acesso em: 27 jul. 2020.

REDDY, B. S.; CHATTERJI, B. N. An fft-based technique for translation, rotation, and scale-invariant image registration. **IEEE Transactions on Image Processing**, v. 5, n. 8, p. 1266–1271, ago. 1996. <<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=506761>>, Acesso em: 24 jan. 2020.

RISTIN, M. et al. Incremental learning of ncm forests for large-scale image classification. **2014 IEEE Conference on Computer Vision and Pattern Recognition**, p. 3654–3661, 2014.

SARVAIYA, J. N.; PATNAIK, S.; KOTHARI, K. Image registration using log polar transform and phase correlation to recover higher scale. **Journal of Pattern Recognition Research**, v. 7, n. 1, p. 90–105, fev. 2012. <<http://www.jprr.org/index.php/jprr/article/view/355/148>>, Acesso em: 19 fev. 2020.

SAXENA, S.; SINGH, R. K. A survey of recent and classical image registration methods. **International Journal of Signal Processing, Image Processing and Pattern Recognition**, v. 7, n. 4, p. 167–176, ago. 2014. <http://article.nadiapub.com/IJSIP/vol7_no4/16.pdf>, Acesso em: 11 fev. 2020.

SCHRUNNER, S. et al. A comparison of supervised approaches for process pattern recognition in analog semiconductor wafer test data. **2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)**, p. 820–823, dez. 2018. <<https://ieeexplore.ieee.org/document/8614156>>, Acesso em: 24 jul. 2020.

SCHUBERT, F.; MIKOLAJCZYK, K. Robust registration and filtering for moving object detection in aerial videos. **2014 22nd International Conference on Pattern Recognition**, p. 2808–2813, 2014.

SCOPUS. Title-abs-key ("sift"). **Scopus**, 2020. <<https://www.scopus.com/search/form.uri?display=basic>>, Acesso em: 26 mar. 2020.

SHARMA, K.; GOYAL, A. Classification based survey of image registration methods. **2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)**, p. 1–7, jul. 2013. <<https://ieeexplore.ieee.org/document/6726741>>, Acesso em: 7 ago. 2020.

SMITH, J. O. **Mathematics of the Discrete Fourier Transform (DFT)**. [S.l.]: W3K Publishing, 2002. <<https://www.chegg.com/textbooks/mathematics-of-the-discrete-fourier-transform-dft-1st-edition-9780974560700-0974560707>>, Acesso em: 08 jun. 2020. ISBN 0974560707.

STONE, H. S.; TAO, B.; MCGUIRE, M. Analysis of image registration noise due to rotationally dependent aliasing. **J. Visual Commun. Image Representation**, v. 14, n. 2, p. 114–135, maio 2003. <[https://doi.org/10.1016/S1047-3203\(03\)00002-6](https://doi.org/10.1016/S1047-3203(03)00002-6)>, Acesso em: 30 abr. 2020.

THEVENAZ, P.; RUTTIMANN, U. E.; UNSER, M. A pyramid approach to subpixel registration based on intensity. **IEEE Transactions on Image Processing**, v. 7, n. 1, p. 27–41, 1998.

TONG, X. et al. Image registration with fourier-based image correlation: A comprehensive review of developments and applications. **IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING**, v. 12, n. 10, p. 4062–4081, out. 2019. <<https://ieeexplore.ieee.org/document/8844710>>, Acesso em: 04 mar. 2020.

WANG, W. et al. A robust affine invariant point extraction algorithm for image registration. **2017 10th International Symposium on Computational Intelligence and Design**, p. 79–82, dez. 2017. <<https://ieeexplore.ieee.org/document/8275724>>, Acesso em: 27 jul. 2020.

WANG, X. et al. A new method of speeded up robust features image registration based on image preprocessing. **2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)**, p. 317–321, jul. 2018. <<https://ieeexplore.ieee.org/document/8666894>>, Acesso em: 7 ago. 2020.

WANG, Z. et al. Hcr-net: A hybrid of classification and regression network for object pose estimation. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18), p. 1014–1020, 2018.

WEBOFSCIENCE. Topic ("sift"). **WebOfScience**, 2020. <http://apps-webofknowledge.ez314.periodicos.capes.gov.br/Search.do?product=WOS&SID=8BQwIByZBkZ4VQiIsvU&search_mode=GeneralSearch&prID=ed3578a6-62bd-4167-9548-22b5733a3e39>, Acesso em: 26 mar. 2020.

YANG, Y.; YUANLI, L. A multi-source image registration algorithm based on combined line and point features. **20th International Conference on Information Fusion**, p. 1–6, jul. 2017. <<https://ieeexplore.ieee.org/document/8009690>>, Acesso em: 27 jul. 2020.

ZHANG, Q.; ZHANG, Z.; ZENG, D. A fast automatic and robust image registration algorithm. **2011 International Conference on Virtual Reality and Visualization**, p. 307–310, 2011. <<https://ieeexplore.ieee.org/document/6092736>>, Acesso em: 31 mar. 2020.

ZHU, H.; LEI, J.; TIAN, X. A pattern recognition system based on computer vision — the method of chinese chess recognition. **2008 IEEE International Conference on Granular Computing**, p. 1–4, ago. 2008. <<https://ieeexplore.ieee.org/document/4664655>>, Acesso em: 01 jun. 2020.

ZHU, H. et al. Sar image registration based on multifeature detection and arborescence network matching. **IEEE Geoscience and Remote Sensing Letters**, v. 13, n. 5, p. 706–710, 2016. <<https://ieeexplore.ieee.org/document/7442791>>, Acesso em: 31 mar. 2020.

ZITOVÁ, B.; FLUSSER, J. Image registration methods: a survey. **Image and Vision Computing**, v. 21, p. 977–1000, out. 2003. <https://www.researchgate.net/profile/Jan_Flusser/publication/222648347_Image_Registration_Methods_A_Survey/links/5ba9f31c92851ca9ed238b48/Image-Registration-Methods-A-Survey.pdf>, Acesso em: 11 fev. 2020.

APÊNDICE A – ANÁLISE DE CORRELAÇÃO E DENSIDADE ESPECTRAL

Uma ferramenta fundamental para a análise de sinais aleatórios é a correlação entre dois sinais. Essa ferramenta provê uma medida de similaridade entre esses dois sinais. A sequência de correlação entre dois sinais de energia discretos $x(n)$ e $y(n)$ é definida por (MANOLAKIS; INGLE; KOGON, 2005) como¹

$$r_{xy}(l) = \sum_{n=-\infty}^{+\infty} x(n)y^*(n-l) \quad (\text{A.1})$$

em que l é chamado de atraso.

A Densidade Espectral de Potência Cruzada (*Cross-Power Spectral Density*), também chamada de Espectro de Potência Cruzada (*Cross-Power Spectrum*), de dois processos estocásticos de **média zero** e conjuntamente estacionários provê uma descrição de suas relações estatísticas no domínio da frequência, sendo definida como a Transformada de Fourier de Tempo Discreto (DTFT – *Discrete Time Fourier Transform*) da correlação cruzada desses processos (MANOLAKIS; INGLE; KOGON, 2005):

$$\tilde{r}_{xy}(e^{j\omega}) = \sum_{l=-\infty}^{+\infty} r_{xy}(l)e^{-j\omega l} \quad (\text{A.2})$$

A correlação cruzada $r_{xy}(l)$ pode ser recuperada pela DTFT inversa, conforme

$$r_{xy}(l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{r}_{xy}(e^{j\omega})e^{j\omega l} d\omega \quad (\text{A.3})$$

O espectro cruzado $\tilde{r}_{xy}(e^{j\omega})$ é, em geral, uma função complexa de ω , sendo (MANOLAKIS; INGLE; KOGON, 2005)

$$\tilde{r}_{xy}(e^{j\omega}) = \tilde{r}_{yx}^*(e^{j\omega}) \quad (\text{A.4})$$

Isso implica que $r_{xy}(e^{j\omega})$ e $r_{yx}(e^{j\omega})$ possuem a mesma magnitude mas fases opostas.

Outra figura de mérito útil na avaliação de relações estatísticas no domínio da frequência é o espectro cruzado normalizado (*normalized cross-spectrum*):

$$\tilde{g}_{xy}(e^{j\omega}) \triangleq \frac{\tilde{r}_{xy}(e^{j\omega})}{\sqrt{\tilde{r}_x(e^{j\omega})}\sqrt{\tilde{r}_y(e^{j\omega})}} \quad (\text{A.5})$$

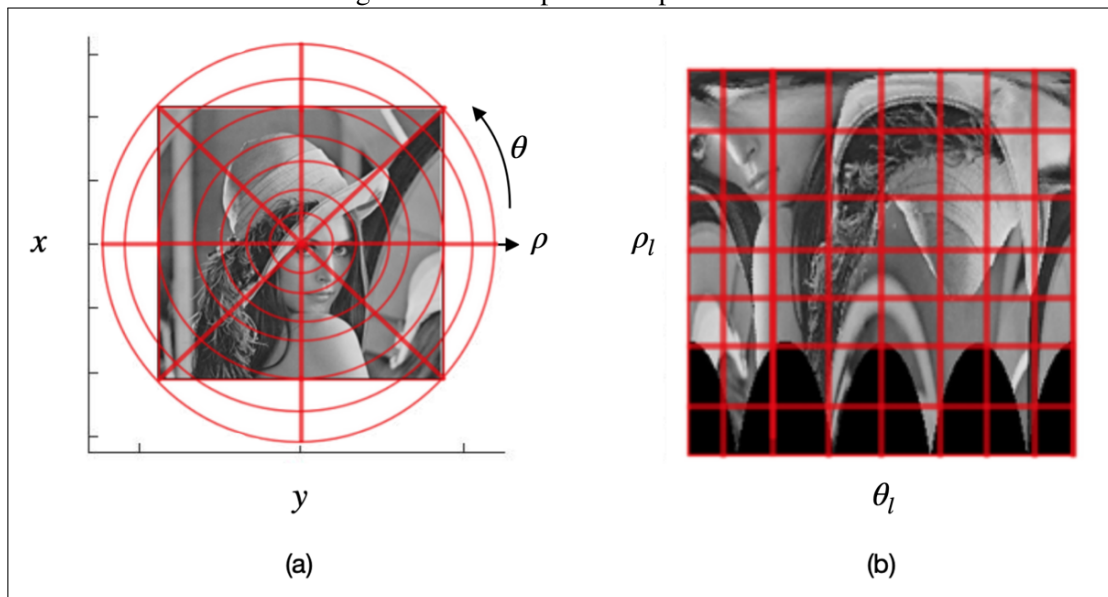
¹Ainda que não esteja apontado em (MANOLAKIS; INGLE; KOGON, 2005), esta definição se trata de uma estimação baseada na hipótese de ergodicidade.

APÊNDICE B – MAPEAMENTO LOG-POLAR

Como abordado no Capítulo 2, Reddy e Chatterji (1996) não fornecem detalhes sobre o desenvolvimento de (2.36) e apresentam ρ como uma constante. Visto que a diferença de escala entre as imagens não é constante, deduz-se que existe um equívoco na equação (provavelmente um erro de digitação na redação do artigo).

Reddy e Chatterji (1996) escolheram um logaritmo de base 1,044, no lugar do logaritmo natural, como base para a conversão de escala. Segundo os autores, essa escolha foi feita com base no tamanho da imagem e com a previsão de que com a base de 1,044 o mapeamento para o plano log-polar de uma imagem com tamanho 256×256 resulta em apenas 128 linhas, ocupando somente metade do tamanho da imagem original na nova imagem. Esse valor pode ser variado conforme o tamanho da imagem e a precisão desejada. O mapeamento de coordenadas é exemplificado com uma imagem na Figura 14.

Figura 14 – Exemplo de Mapeamento



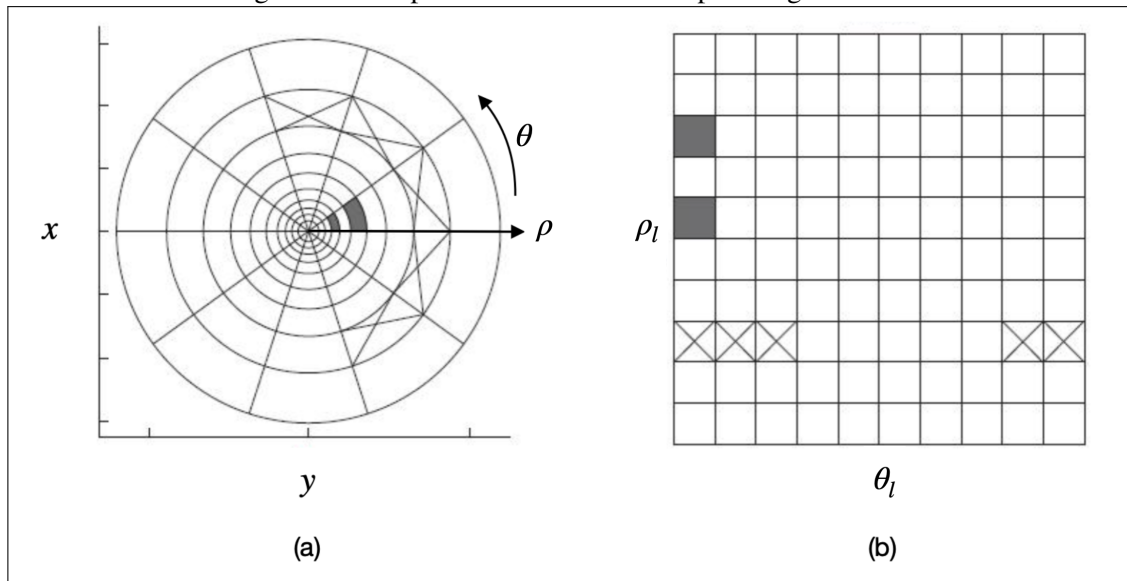
(a) Imagem no plano cartesiano e (b) Imagem no plano log-polar. Fonte: Autor(2021).

O mapeamento entre o plano cartesiano e o plano log-polar descrito em (2.31) está ilustrado na Figura 15. Na Figura 15(a) é mostrado o plano original da imagem ($x-y$) e na Figura 15(b) é mostrado o plano log-polar, em que

$$\rho_l = \ln(\rho) \quad (\text{B.1})$$

A variável ρ descreve a distância entre o pixel e o centro das circunferências descritas na Figura 15(a). Essas circunferências, que compõem anéis de tamanhos crescentes a partir do centro, quando comparadas com as regiões uniformes na figura à direita, ilustram o caráter logarítmico do mapeamento, que pode ser observado pelo posicionamento entre as áreas hachuradas e marcadas com um “×” em ambas as figuras.

Figura 15 – Mapeamento de Cartesiano para Log-Polar



(a) Plano cartesiano e (b) Plano log-polar. Fonte: Adaptado de (SARVAIYA; PATNAIK; KOTHARI, 2012).

Considerando-se as configurações escolhidas por Reddy e Chatterji (1996), é possível mostrar que o mapeamento inverso do sistema de referência da Figura 15(b) para a Figura 15(a) resulta em

$$\rho = 1,044^{\rho_l} \quad (\text{B.2})$$