

Universidade Federal de Minas Gerais  
Escola de Engenharia  
Curso de Graduação em Engenharia de Controle e Automação

## **Implementação de um sistema de leitura automática de documentos**

Ana Claudia Abascal Gobetti

Orientador: Prof. Frederico Gadelha Guimarães, Dr.  
Supervisor: Eng. Ludmila de Oliveira Moreira Lopes

Belo Horizonte, Julho de 2017



## **Monografia**

### **Implementação de um sistema de leitura automática de documentos**

Monografia submetida à banca examinadora designada pelo Colegiado Didático do Curso de Graduação em Engenharia de Controle e Automação da Universidade Federal de Minas Gerais, como parte dos requisitos para aprovação na disciplina Projeto Final de Curso II.

Belo Horizonte, Julho de 2017

# Resumo

Esse projeto tem como objetivo desenvolver um sistema que permita a leitura automática de documentos pessoais. Para isso, serão utilizadas técnicas de visão computacional e reconhecimento ótico de Caracteres (OCR). Dessa forma, o sistema desenvolvido tem como objetivo não somente digitalizar e ler os documentos, mas identificar as áreas de interesse, nas quais estão contidos os dados e ele deve ser flexível quando ao meio de entrada da imagem do documento e permitir imagens scaneadas com diferentes resoluções e configurações de contraste e brilho. Destaca-se a importância deste sistema na automação comercial durante o processo de cadastramento de clientes, por exemplo, durante os check-in de hotéis, uma vez que a digitalização e interpretação dos dados contidos no documento de identificação por um computador é feita de forma muito mais rápida e é menos propensa a erros do que a leitura feita por um operador, que está sujeito à diversos fatores que podem influenciar na sua eficiência.

Dessa forma, percebe-se que a existência de ferramentas que não só convertam arquivos do formato físico para o formato digital, mas que também leiam e interpretem os dados seria de grande utilidade, uma vez que simplificaria o trabalho, reduzindo o tempo de espera em diversas atividades, como nos casos em que é necessário realizar um cadastro pessoal, como ao alugar um quarto de hotel ou alugar um carro ou equipamento.



# **Lista de Figuras**

2.1	Comparação entre os dois modelos de carteira de habilitação existentes no Brasil no ano de 2017 . . . . .	5
2.2	Categorias de imagens, coloridas, escala de cinza e binárias . . . . .	6
2.3	Etapas do processamento de imagens . . . . .	7
2.4	Modelo de scanner utilizado em sistemas de digitalização de documentos existente atualmente . . . . .	10
3.1	Metodologia adotada para identificação dos campos. Na figura o ponto (x,y) determina a coordenada inicial do retângulo onde o dado está contido, L a largura do campo e A a altura . . . . .	13
3.2	Arquitetura do sistema de leitura automática . . . . .	14
3.3	Diagrama de funcionamento do sistema . . . . .	15
4.1	a) Histograma da imagem do documento após a conversão para escala de cinza e b) Histograma após a seleção da faixa cinza de interesse . . . . .	18
4.2	a) Imagem da CNH convertida em escala de cinza, sem tratamento e b) após a seleção da faixa cinza de interesse . . . . .	19
4.3	Resultado do processamento após a obtenção das linhas . . . . .	20
4.4	Interface desenvolvida para o protótipo . . . . .	21
4.5	Resultados obtidos para o novo modelo de CNH . . . . .	22

# Capítulo 1

## Introdução

O processo cognitivo-visual do ser humano é um assunto extremamente complexo, uma vez que a visão não se resume somente à formação de uma imagem do ambiente que nos rodeia, mas envolve também análise, categorização e reconhecimento dos componentes que constituem tal imagem, bem como interações com outras funções cognitivas, como emoções, linguagem, memória, entre outros.

A simulação do processo citado acima por meio de computador caracteriza os sistemas de visão computacional. A visão computacional pode ser definida como o conjunto de métodos e técnicas que tornam os sistemas capazes de extrair e interpretar as informações presentes em uma imagem. Um dos seus maiores objetivos é a busca por um modelo de representação genérico que se aproxime ao processo realizado por um ser humano.

Outra aplicação destes sistemas surgiu da necessidade do aumento da fiabilidade das informações obtidas. Isto se deve ao fato de que o processo de classificação de imagens pelo homem é muito eficaz, porém é sujeito a falhas que podem ser ocasionadas pelo cansaço, fadiga, dentre outros fatores. Dessa forma, a utilização de visão computacional não substitui o homem em suas tarefas, porém pode auxiliá-lo a diminuir erros.

Dessa forma, percebe-se que estes sistemas podem ser utilizados em diversas aplicações nos mais variados domínios, como por exemplo: medicina, automação industrial, automação comercial, sensoriamento remoto, etc.

### 1.1 Motivação e Justificativa

A evolução da tecnologia permitiu o aumento da velocidade de processamento dos computadores, possibilitando a realização de diversas tarefas que não eram possíveis até pouco tempo atrás, gerando capacidade para realização de diversas novas tarefas. Ao mesmo tempo, o avanço da tecnologia fez com que fosse crescente a necessidade de se acessar dados de forma mais rápida e eficiente. Além disso, existe a necessidade de arquivar e gerir grandes quantidades de informação. Nos dias atuais, a tendência é a utilização de Sistemas de Informação para armazenar esses dados, facilitando o seu acesso, gestão e utilização.

No entanto, algumas informações ainda são armazenadas no formato físico, como é o caso dos documentos pessoais de identificação, que podem ser digitalizados. Este dualismo requer que, muitas vezes, seja necessário transferir informações do meio físico para o meio

digital, como em alguns sistemas em que pode ser necessário cadastrar um cliente, digitando o seu nome, número do documento, data e local de nascimento, entre outras informações.

Porém não existe atualmente uma maneira rápida e automática de realizar esta conversão dos dados, sendo necessário um procedimento manual de interpretação da informação contida no documento. Este é um processo que, além de cansativo e repetitivo, é um grande consumidor de recursos humanos e está sujeito à falhas. Dessa forma, percebe-se que a existência de ferramentas que não só convertam arquivos do formato físico para o formato digital, mas que também leiam e interpretem os dados seria de grande utilidade, uma vez que simplificaria o trabalho, reduzindo o tempo de espera em diversas atividades, como nos casos em que é necessário realizar um cadastro pessoal, como ao alugar um quarto de hotel ou alugar um carro ou equipamento.

## 1.2 Objetivos do Projeto

O objetivo deste projeto é desenvolver um software que utilize de conceitos de visão computacional capaz de extrair informações pessoais de um documento de identificação, no caso a Carteira Nacional de Habilitação (CNH).

## 1.3 Organização do Trabalho

Este trabalho está organizado da seguinte forma:

- Introdução: Apresenta os sistemas de visão computacional, bem como suas aplicações, vantagens e limitações tecnológicas envolvidas. Além disto também são apresentados os objetivos e a organização do trabalho.
- Revisão Bibliográfica: Neste capítulo será apresentada a fundamentação teórica necessária para o completo entendimento do projeto. Serão abordados temas como as principais características de uma carteira de habilitação, visão computacional, processamento de imagens, entre outros.
- Materiais e Métodos: Nesta seção serão descritos todos os passos necessários para implementar o sistema leitura automática da CNH. E serão definidas as formas de avaliação do sistema.
- Resultados: Serão apresentados os resultados do sistema implementado no capítulo anterior.
- Conclusão: Serão apresentadas a conclusões bem como proposições de trabalhos futuros.

# Capítulo 2

## Revisão Bibliográfica

Neste capítulo serão abordados todos os conhecimentos necessários para o desenvolvimento do projeto, como as características da Carteira Nacional de Habilitação, as bibliotecas e linguagens utilizadas para o desenvolvimento e principais algoritmos e métodos usados na implementação do sistema de leitura automática.

### 2.1 Carteira Nacional de Habilitação

A CNH, Carteira Nacional de Habilitação, que também é conhecida como carteira de motorista é um documento de identificação obrigatório para qualquer cidadão que pretenda conduzir um veículo automotor. Atualmente o código brasileiro divide a CNH em cinco categorias de acordo com o tipo de veículo que o condutor está habilitado a conduzir, segundo [DETTRAN PR \[2017\]](#):

- A:** condutor de veículo motorizado de duas ou três rodas, com ou sem carro lateral (motos);
- B:** condutor de veículo motorizado não abrangido pela categoria A, com peso bruto total inferior a 3.500 quilogramas e lotação máxima de oito lugares, além do motorista (automóveis);
- C:** condutor de veículo motorizado usado para transporte de carga, com peso bruto superior a 3.500 quilogramas (como caminhões);
- D:** condutor de veículo motorizado usado no transporte de passageiros, com lotação superior a oito lugares além do motorista (ônibus e vans, por exemplo);
- E:** condutor de combinação de veículos em que a unidade conduzida se enquadre nas categorias B, C ou D e cuja a unidade acoplada ou rebocada tenha peso bruto de 6 mil quilogramas ou mais; ou cuja lotação seja superior a oito lugares; ou, ainda, que seja enquadrado na categoria trailer;

A primeira CNH só pode ser retirada nas categorias A ou B por meio de cursos teóricos preparatórios, médico e psicotécnico. Após a primeira habilitação existem algumas regras para mudança de categoria ([DETTRAN MG \[2017\]](#)):

- Categoria A e B: ter mais de 18 anos completos;
- Categoria C: ter, no mínimo, um ano na categoria "B";
- Categoria D: ter 21 anos completos, estar habilitado no mínimo a 2 anos na categoria B ou 1 ano na categoria "C";
- Categoria E: ter 21 anos completos, estar habilitado, há um ano nas categorias "C" ou "D";

Atualmente a CNH possui, além dos dados acerca da habilitação, fotografia, número da carteira de identidade (RG) e do Cadastro de Pessoa Física (CPF). Assim a CNH pode ser utilizada como um documento de identificação pessoal em todo território nacional (Art. 159 do [CTB \[1997\]](#)). Além disto a CNH tem data de validade, ou seja, motoristas com menos de 65 anos devem renovar a carteira a cada 5 anos e após os 65 anos ou para alguns casos especiais o prazo de renovação é de 3 anos. Dessa forma, garante-se que o documento apresentado é sempre recente.

Por servir como documento de habilitação e documento de identificação, este é um dos documentos mais falsificados do Brasil. Segundo a [Serasa](#), a cada 15 segundos alguém tenta cometer uma fraude usando documentos falsos no Brasil. Estes documentos falsificados são utilizados na tentativas de golpes como abrir contas bancárias em nome de outras pessoas, pedir financiamentos, cartões de crédito entre outros. Com o intuito de impedir a falsificação e adulteração do documento de habilitação, o Conselho Nacional de Transito (Contran) busca investir em novos dispositivos de segurança. Dessa forma, a partir do dia 31 de dezembro de 2017 entra em vigor um novo modelo de CNH, conforme pode ser visto na figura 2.1.

Pela análise da imagem ?, percebe-se que há uma grande alteração de tonalidades no novo modelo, mas os campos continuam seguindo a mesma ordem, desta forma as alterações realizadas não tem muito impacto no desenvolvimento deste trabalho. Além da clara mudança de cor, haverá um mapa do estado, inclusão de dois números de identificação nacional (Registro Nacional e número de espelho da CNH), entre outros dispositivos de segurança. A mudança para o novo modelo é gradual, portanto os motoristas não precisarão trocar o documento atual caso ele ainda seja válido, dessa forma os dois modelos irão coexistir por 5 anos.

## 2.2 Imagem Digital

As imagens digitais podem ser produzidas por meio de diversos aparelhos, como câmeras de vídeo, câmeras fotográficas, scanners, aparelhos de raio-X, entre outros. A tecnologia permite a codificação digital de documentos analógicos, em forma de imagem digital para armazenamento e transmissão em sistemas computadorizados.

Uma imagem digital é a representação de uma figura bidimensional como uma sequência finita de valores digitais. Cada um destes valores representa um pixel ou uma componente de cor de um pixel, no caso de imagens coloridas. Um pixel é o menor elemento de uma imagem digital, a partir deles é possível medir a qualidade da imagem, sendo comum atribuir o termo resolução para se referir ao número de pixels em altura e largura em uma foto.

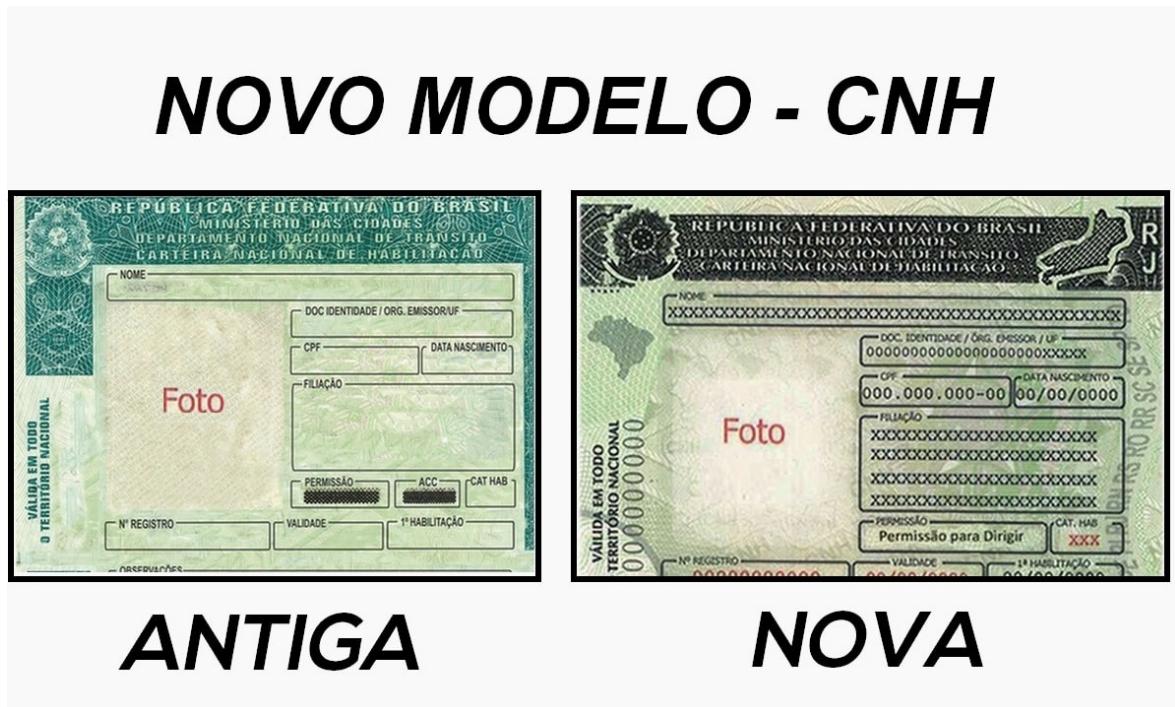


Figura 2.1: Comparação entre os dois modelos de carteira de habilitação existentes no Brasil no ano de 2017

O valor de cada pixel é obtido pela quantização do sinal analógico (carga elétrica) obtido pelos dispositivos de captura, citados anteriormente. Os três tipos principais de imagens são:

- Imagens coloridas: cada pixel é representado por pelo menos três informações, uma para cada canal de cor, por exemplo, o RGB que será um dos itens discutidos nas próximas seções. O valor de cada canal e a composição da imagem colorida depende do espaço de cores utilizado.
- Imagens em escala de cinza: Cada pixel é representado por apenas um valor, que pode variar de 0, que indica a cor preta, até 255, que indica a cor branca, totalizando 256 tons de cinza diferentes.
- Imagens Binárias: Também conhecida como preta e branca. Nestas imagens cada pixel pode assumir valores 0 ou 1. Normalmente surge no processamento de uma imagem digital como resultado de uma máscara, ou certas operações como segmentação, threshold e ruído (estes conceitos serão discutidos nas próximas sessões).



Figura 2.2: Categorias de imagens, coloridas, escala de cinza e binárias

## 2.3 Visão Computacional

A visão computacional é um campo que inclui métodos para a aquisição, processamento, análise e compreensão de imagens. A finalidade dessas análises e processamentos é a obtenção de informações numéricas ou simbólicas de imagens, por exemplo, na forma de decisões ([SHAPIRO \[2001\]](#)). Dessa forma, o objetivo de um sistema de visão computacional é extrair informações úteis a partir de dados de imagem usando modelos construídos com a ajuda da geometria, física, estatística e teoria de aprendizagem.

O sistema cognitivo do ser humano processa e extrai informações de imagens a todo momento, e utilizar de sistemas computacionais para realizar a mesma atividade é algo extremamente complexo e que exige diversos conhecimentos e abstrações. Para isso a organização de cada sistema depende muito da sua aplicação. No entanto, existem algumas funções típicas que são encontradas em muitos sistemas de visão computacional, normalmente são executadas conforme a figura 2.3:

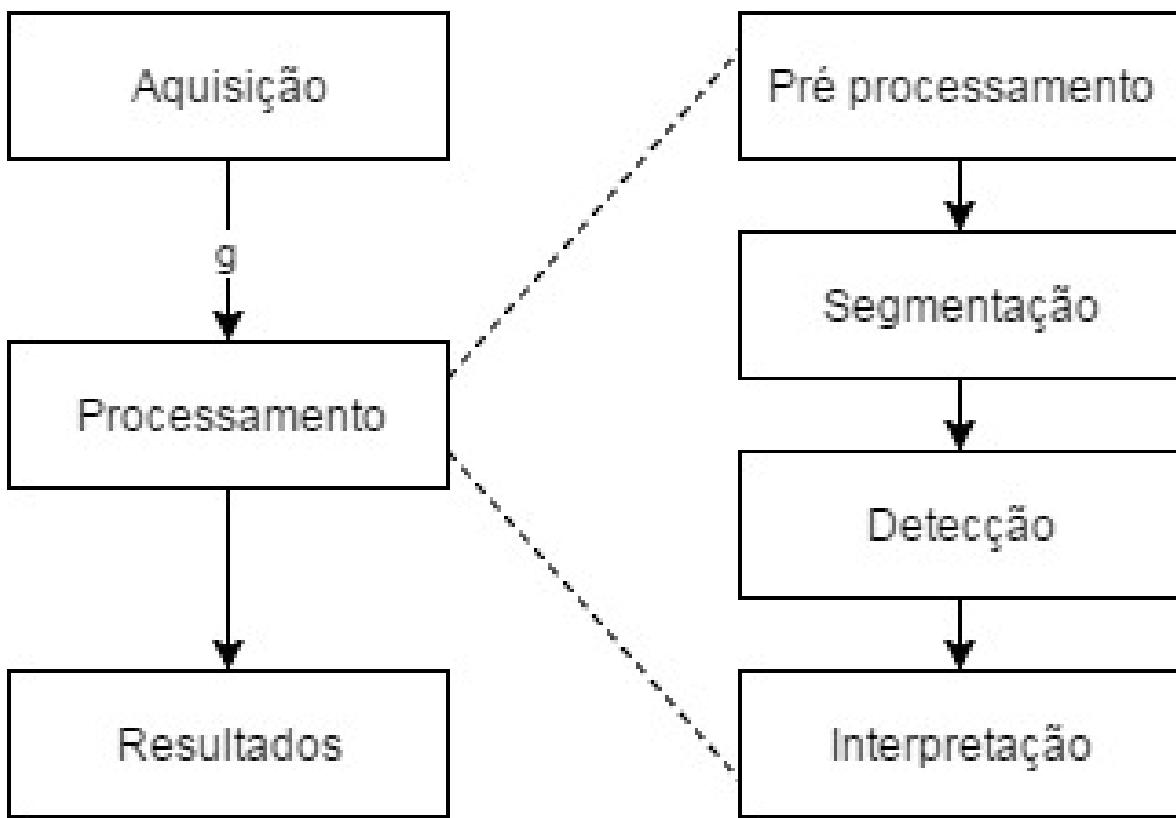


Figura 2.3: Etapas do processamento de imagens

### Aquisição de imagens

A primeira etapa consiste em adquirir a imagem digital a partir de uma câmera digital, webcam ou scanner.

### Processamento

Uma imagem digital pode possuir muitas informações que podem influenciar negativamente na tomada de decisão do sistema. O processamento permite que estas informações indesejadas sejam ignoradas ou retiradas (ZUECH [1988]). O processamento de imagens envolve técnicas de transformação visando melhorar, segmentar e extrair as características desejadas da imagem. As principais etapas deste processo são:

- Pré-processamento: Nessa etapa a imagem é submetida a um processo de realce dos contrastes, deixando-a mais nítida para facilitar as etapas seguintes.
- Segmentação: É uma etapa em que serão delimitados os objetos de interesse.
- Detecção: Após a segmentação é necessário determinar quais das áreas delimitadas são realmente necessárias para extrair a informação desejada, no sistema em questão, será possível determinar e classificar o dado da CNH.
- Interpretação: Interpretação e classificação dos dados detectados.

Nas subseções seguintes serão descritas algumas técnicas e métodos utilizados nesta aplicação.

### 2.3.1 Conversão para Escala de Cinza

Como já foi discutido nas seções anteriores, as imagens digitais possuem diversas representações. A mais comum é a conhecida como RGB, que é uma representação formada por 3 canais: vermelho (Red), verde (Green) e azul (Blue), nela são produzidas várias cores por meio da combinação destas três cores básicas. Nessa representação cada pixel possui três bytes, um para cada cor, que podem possuir valores entre 0 e 255. Para reproduzir a cor preta as três dimensões devem possuir valores iguais a 0. Para obter a cor vermelha basta selecionar 255 na componente R do pixel e 0 nas demais componentes. Dessa forma, é possível criar a cor desejada a partir da combinação ponderada de cada um dos componentes. Porém algumas operações do processamento de imagens são facilitados quando os pixels têm apenas uma dimensão, ou seja, quando a imagem está em escala de cinza ou binária. Para converter uma imagem colorida em escala de cinza é feita uma média simples entre os valores de RGB do pixel e esta média é atribuída ao novo valor.

### 2.3.2 Binarização

A identificação de partes de interesse em uma imagem é uma das etapas mais críticas no processamento de imagens. A binarização de uma imagem consiste em dividir a imagem em regiões de interesse, isolando aqueles pixels que não fazem parte do mesmo.

Os métodos mais simples de limiarização utilizam um único ponto de corte também conhecido por threshold. O threshold é um valor limite utilizado em um critério de seleção. Todos os pixels de uma imagem são comparados a esse critério e são alterados conforme a necessidade. Um exemplo da aplicação de threshold é comparar todos os pixels a um valor limite, caso sejam maiores ou iguais a esse limiar são transformados em pretos (valor 0), e transformados em brancos (valor 255), caso contrário.

### 2.3.3 Segmentação baseada em bordas

Na detecção de bordas são analisadas as descontinuidades nos níveis de cinza. Neste trabalho ela é empregada para delimitar os campos dos dados no documento.

Uma borda é o limite entre duas regiões com propriedades relativamente distintas de nível de cinza. As bordas na imagem caracterizam os contornos dos objetos e são bastante úteis para segmentação e identificação dos pontos de interesse. Porém, quando a imagem é diferenciada, todas as variações dos níveis de cinza são detectadas e, por consequência, detectam-se também bordas indesejadas. Para que as bordas indesejadas, provenientes de ruído ou textura da imagem, não sejam detectadas, deve-se tratar a imagem a fim de suavizá-la. Contudo, existem efeitos negativos ligados à suavização, como a perda de informação e o deslocamento de estruturas de feições relevantes na imagem.

Historicamente, uma variedade de detectores de bordas tem sido desenvolvida visando diferentes propósitos, com formulações matemáticas diferenciadas e com propriedades

algorítmicas distintas. Em 1986, Canny desenvolveu um processo de detecção de bordas a partir de critérios de quantificação de desempenho de operadores de bordas conhecidos como os critérios de detecção e de localização.

### 2.3.4 Transformada de Hough

A transformada é uma importante técnica para extrair características de uma imagem. Foi desenvolvida por Paul Hough em 1962 e se tornou popular no processamento de imagens em 1981, com a publicação do artigo "Generalizing the Hough transform to detect arbitrary shapes". A proposta desta técnica é basicamente a identificação de qualquer forma que possa ser representada matematicamente, porém seu uso mais comum é para detecção de linhas, círculos e eclipses em imagens binárias. Normalmente a imagem em que a transformada é aplicada passou anteriormente pela operação de reconhecimento de bordas.

## 2.4 Sistemas OCR

Segundo [Bhaskar et al. \[2010\]](#), o algoritmo de reconhecimento ótico de caracteres (Optical Character Recognition, OCR) é a tradução, eletrônica ou mecânica, de imagens escaneadas de documentos escritos à mão, datilografados em máquinas de escrever ou impressos, para textos codificados em máquinas. As aplicações dos sistemas OCR são as mais variadas, indo desde o reconhecimento de placas de automóveis ou de outros tipos de placas e avisos, passando pela digitalização de documentos para o formato texto, até o reconhecimento de caligrafias manuais.

Segundo [Mori et al. \[1992\]](#) os primeiros avanços dessas tecnologias começaram nos anos 50, inicialmente apenas conceitualmente, mas gradualmente avançando para ferramentas mais concretas. As primeiras versões eram bastante sujeitas a falhas e se baseavam em comparações de documentos ou em imagens como modelos. Com o passar do tempo, concluiu-se que a comparação de modelos não era suficiente para obter bons resultados e começaram os estudos para utilização da análise estrutural dos caracteres para sua identificação. Essa mudança de abordagem deu origem aos sistemas atuais.

Atualmente, com o barateamento dos computadores, aumento da capacidade de processamento ([Lins \[2012\]](#)) e do grau de maturidade da tecnologia OCR, os sistemas de leitura de documentos estão cada vez mais acessíveis às pessoas e novos algoritmos, bem como técnicas para pré-processamento, extração e classificação de dados foram criados ([Cheriet \[2007\]](#)).

Muito além de reconhecer texto em imagens, leitura automática de documentos refere-se a todo e qualquer processo ou mecanismo que permita extrair, de forma automatizada, a informação textual existente nas imagens dos documentos. Neste projeto pretende-se não somente extraír o texto dos documentos, mas também atribuir sentido à essa informação.

Existem alguns sistemas de leitura automatizada de documentos, mas a maioria deles utiliza um scanner próprio, como o que pode ser visto na figura 2.4, para obter a imagem de entrada. Dessa forma, o sistema fica limitado e restrito. Alguns fornecedores, como Mitek e TrustID, possuem ferramentas mobile e web para leitura automática de documentos, porém

o fato de não serem soluções nacionais torna o suporte um pouco mais lento e as atualizações quanto à legislação e validade de documentos mais demorada.



Figura 2.4: Modelo de scanner utilizado em sistemas de digitalização de documentos existente atualmente

## 2.5 Resumo do Capítulo

Esse capítulo tem como objetivo uma revisão bibliográfica da teoria de visão computacional e etapas de processamento de imagens contextualizadas no sistema de leitura automática de documento. Com base nas informações apresentadas, no próximo capítulo será detalhado como essas técnicas foram utilizadas para a elaboração do projeto realizado.

# Capítulo 3

## Materiais e Métodos

### 3.1 Ambiente de Desenvolvimento

Uma linguagem de programação é uma linguagem artificial projetada para comunicar instruções a uma máquina, especialmente um computador. Linguagens de programação podem ser utilizadas para criar os programas que controlam o comportamento de uma máquina.

A descrição de uma linguagem de programação é geralmente dividida em dois componentes da sintaxe (forma) e semântica (significado). Alguns idiomas são definidos por um documento de especificação, como por exemplo, a linguagem de programação C é especificada por um padrão ISO. ([ISO/IEC \[2011\]](#)).

#### 3.1.1 Linguagem C#

C# é uma linguagem de programação orientada a objeto desenvolvida pela Microsoft em meados de 1999 com base na linguagem C++ que permite criar uma grande variedade de aplicativos seguros e robustos que são executados no .NET Framework. A intenção da Microsoft foi criar uma linguagem de uso geral simples, robusta, orientada objetos e fortemente tipada. É possível usar C# para criar aplicativos cliente do Windows, Web Services, aplicativos cliente-servidor, entre outros.

#### 3.1.2 Biblioteca OpenCV e EmguCV

OpenCV (Open Source Computer Vision Library) é uma biblioteca livre ao uso acadêmico e comercial, para o desenvolvimento em linguagem C e C++, de aplicativos na área de visão computacional ([Bradski and Kaehler \[2008\]](#)). Esta biblioteca possui mais de 2500 algoritmos otimizados, desde os mais simples até os mais modernos, tais como os de Machine-Learning.

O OpenCV pode ter ser utilizado no desenvolvimento de aplicativos com as mais diversas aplicações, desde programas simples como colagem de imagens até programas complexos como auxílio na navegação robótica.

Segundo [Miguel \[2013\]](#)

O OpenCV foi projetado especialmente para eficiência computacional e têm enorme foco em aplicações em tempo real, que utilizam processamento de visão por computador. Foi desenvolvido em C/C++ otimizado e permite tirar partido de processamento multi-core. Confere, ainda, um enorme grau de abstração da programação que requer este tipo de processamento.

EmguCV tem como principal função adaptar o código na biblioteca OpenCV para que possa ser utilizado em plataformas e linguagens compatíveis com o .NET Framework, como C#, VB, VC++, entre outros. Dessa forma, o EmguCV permite a implementação de funcionalidades do OpenCV através do Visual Studio em linguagens de programação como o C#.

### 3.1.3 Tesseract

O Tesseract é a biblioteca opensource responsável pelo reconhecimento ótico dos caracteres, desenvolvida pela HP entre 1985 e 1995 e a partir de 2006 o projeto foi continuado pela Google. Atualmente o Tesseract é considerado a melhor ferramenta OCR opensource ([Bhaskar et al. \[2010\]](#)).

### 3.1.4 Microsoft Visual Studio

Visual Studio é o ambiente de desenvolvimento (IDE) da Microsoft para construção de aplicações em C#, Visual Basic, Visual C#, C++, JavaScript, entre outras linguagens. Com esta ferramenta é possível criar as mais diversas aplicações desktop, aplicativos móveis, serviços Web, dentre outros. A versão do Visual Studio utilizada para o desenvolvimento deste trabalho é a 2015 com .NET Framework 4.

## 3.2 Modelagem do sistema

### 3.2.1 Conceito

A leitura automática de documentos consiste na aquisição e interpretação da informação contida no formato físico. Para este processo são utilizadas tecnologias para digitalizar os documentos, tais como câmeras e scanners, e software para o reconhecimento de caracteres, o OCR. Dessa forma, ao se scanear um documento, será possível não somente a transformação para o formato digital como também obter os dados para o preenchimento de um cadastro pessoal em um sistema de informação de uma empresa, por exemplo.

Dessa forma, percebe-se a importância e utilidade desses sistemas de leitura automática de documentos, uma vez que reduz o trabalho manual para interpretar e digitar os dados do documento, reduzindo o tempo e os custos referentes a estas atividades.

Para interpretação dos dados é necessário definir um modelo de identificação do documento. Neste trabalho, será utilizada a localização das regiões ou segmentos de interesse

### 3.3. LEITURA AUTOMÁTICA DE DOCUMENTO

13

para atribuir sentido ao dado lido. Por exemplo, para a leitura do nome completo é necessário definir as coordenadas (x,y), a largura (L) e altura (A) do campo no documento, como pode ser visto na figura 3.1.

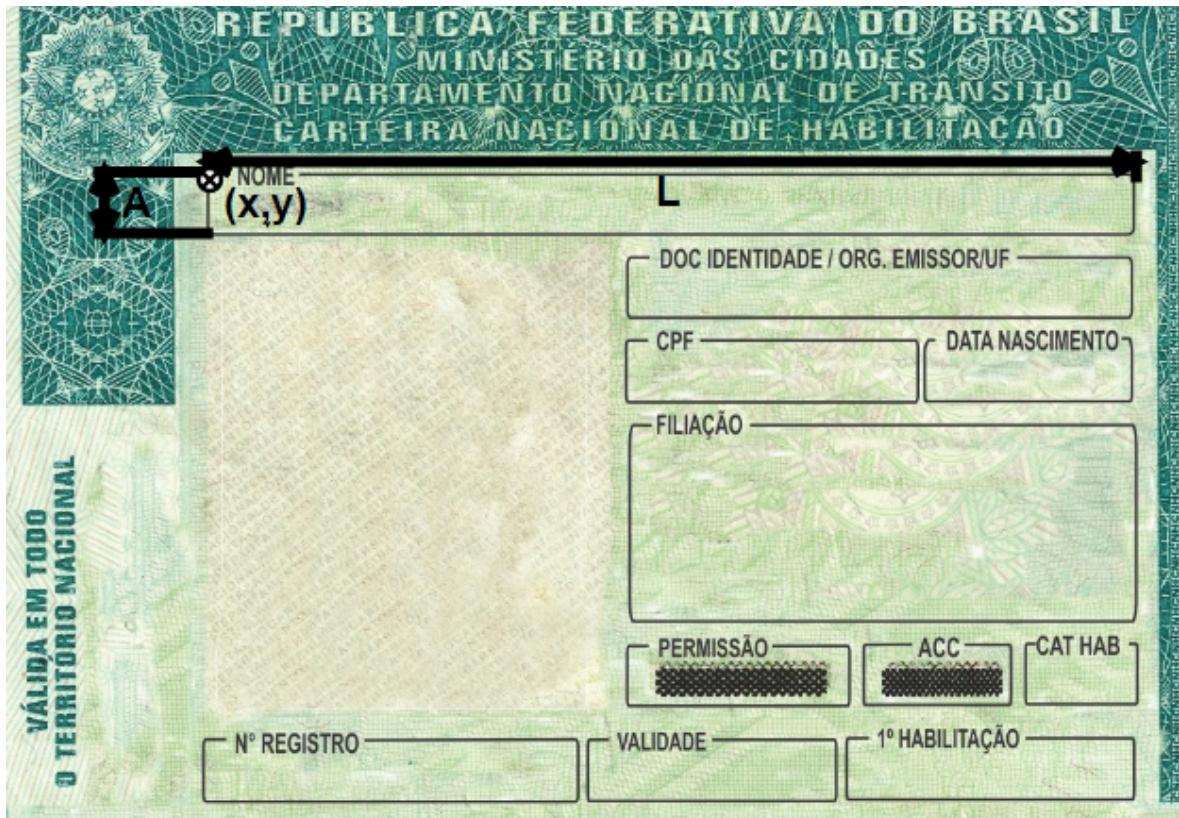


Figura 3.1: Metodologia adotada para identificação dos campos. Na figura o ponto (x,y) determina a coordenada inicial do retângulo onde o dado está contido, L a largura do campo e A a altura

Outro fator importante para este sistema é a independência em relação à digitalização e arquivamento dos dados. Dessa forma, é possível alterar o design da tela de cadastro, ou a forma de armazenamento dos dados sem que seja necessária uma atualização do sistema de leitura automática. Ou seja, a leitura deve acontecer de forma que o sistema a leitura é transparente para o desenvolvedor do SI, como pode ser visto na figura 3.2.

## 3.3 Leitura Automática de Documento

### 3.3.1 Comunicação entre processos

Como foi explicado na seção 3.2.1, o sistema foi projetado para ser independente da interface gráfica e do modelo de dados do usuário. Dessa forma, o protótipo do sistema desenvolvido para este trabalho é constituído de dois projetos executáveis. O primeiro executável será

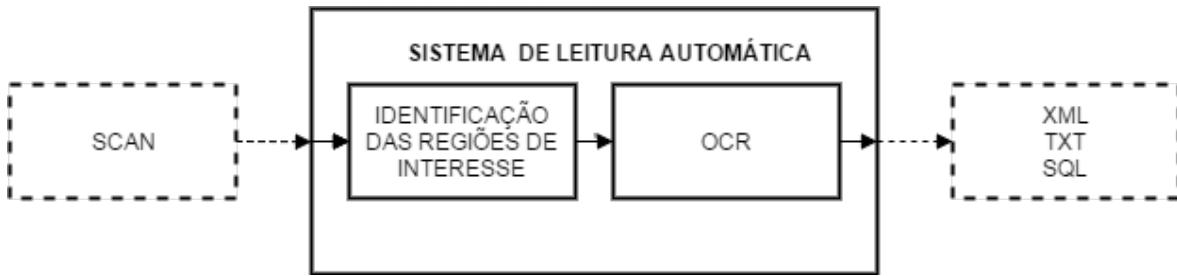


Figura 3.2: Arquitetura do sistema de leitura automática

responsável pela entrada de dados do sistema e receberá o resultado do processamento do documento, exibindo os dados em uma interface gráfica e com a possibilidade de salvar os dados em um arquivo XML. Como foi explicado anteriormente, o projeto foi desenvolvido desta forma para deixar o sistema de leitura automática de dados independente de interfaces gráficas e modelo de dados, permitindo a personalização. O segundo executável é o sistema de leitura de dados em si. Para a comunicação entre os processos foi utilizado o protocolo de comunicação HTTP. O sistema de leitura automática é um servidor HTTP, espera uma requisição em uma porta, com a imagem da CNH passada como parâmetro. Em seguida é realizado o processamento da imagem e o resultado é devolvido na forma de json com os dados lidos.

### 3.3.2 Tratamento da imagem e Leitura do Documento

Como pode ser visto na seção 2.3 um dos primeiros passos de sistemas de visão computacional é o processamento da imagem. No caso deste sistema, é recebida uma imagem colorida do documento da CNH como entrada. O primeiro tratamento é a conversão da mesma para escala de cinza. Em seguida, são iniciadas duas threads de processamento de imagem.

A primeira recebe a imagem e tenta reconhecer um rosto humano no documento. Caso não seja encontrado, o sistema automaticamente invalida o mesmo, uma vez que como pode ser visto na imagem 2.1 a CNH possui uma foto de rosto e não encontrar esta face significa que ou a imagem recebida realmente não é do documento ou não possui qualidade suficiente para a identificação. Para a identificação de faces foi utilizado um método chamado Haar Trainning. Neste método é utilizado um arquivo XML, chamado Haar Classifier, que contém as informações do objeto que se deseja identificar, no caso uma face humana. No caso deste sistema, foi utilizado o arquivo disponibilizado pelo EmguCV.

Na segunda thread é feito um processamento a fim de segmentar a imagem nas regiões de interesse. Inicialmente são detectadas as bordas da imagem, em seguida é utilizado a transformada de Hough para as linhas da imagem, para segmentar as áreas de dados do documento, que possui o contorno delimitado por uma linha escura. Outra abordagem possível para este problema seria identificar os contornos da imagem, identificando a área desejada. Porém esta abordagem se mostrou menos efetiva uma vez que devido à irregularidade da imagem, que possui a descrição do campo, não foi possível identificar

os contornos desejados corretamente. Dessa forma optou-se pela abordagem das linhas.

O sistema aguarda o resultado destas duas threads e após obtê-lo é iniciada a busca pelas regiões de interesse. Essa busca se baseia na localização dos campos em relação à face. Ou seja, inicialmente, busca-se o nome, que é o primeiro campo detectado acima da face. Após identificar as linhas que compõe o campo, é possível extrair as informações do ponto inicial do campo (x,y), do comprimento do campo e da altura, com base nas linhas que se interceptam em um ângulo de 90°. Uma vez identificado estas componentes, é extraída a região de interesse da imagem e esta passa pelo reconhecimento de caracteres.

Foram iniciados três instâncias do leitor OCR, com três dicionários possíveis. Este tratamento foi realizado para reduzir a possibilidade de erros do sistema, uma vez que cada campo possui uma determinada característica, por exemplo, o nome pessoal possui letras de A à Z enquanto campos como o CPF e datas só possuem números e alguns pontos .

Após a leitura de todos os dados, é montado um objeto e ele é retornado na requisição HTTP, que foi comentada na seção 2.2.1. No diagrama da figura 3.3 é possível ver um esquema do funcionamento do sistema.

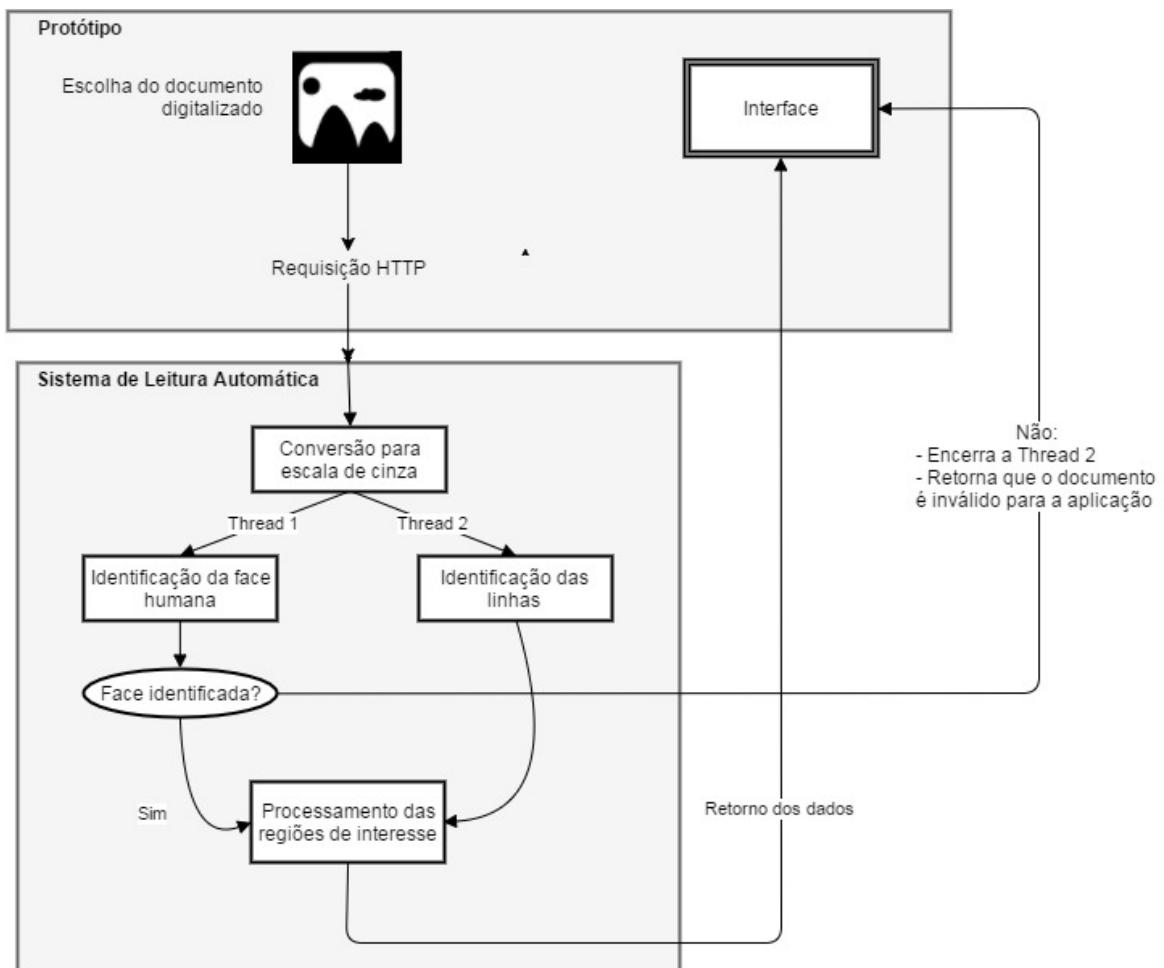


Figura 3.3: Diagrama de funcionamento do sistema



# Capítulo 4

## Resultados

Para a execução do projeto, algumas etapas de desenvolvimento tiveram de ser seguidas: familiarização com a biblioteca, estudo dos módulos envolvidos, leitura dos requisitos, elaboração de documento descrevendo todo o processo de implementação e relacionamento com os diversos módulos, implementação e testes. Neste capítulo serão apresentados os resultados de cada operação realizada no processamento de imagem, que foi detalhado no capítulo anterior.

### 4.1 Leitura automática de documentos

#### 4.1.1 Calibração automática do sistema

Como foi descrito no capítulo anterior, converte-se a imagem para o domínio de escala de cinza, é feito um redimensionamento da mesma e esta será a imagem utilizada como base para todo o sistema de leitura automática. Para segmentação da imagem, um dado importante é a análise do histograma da imagem gerada. O histograma é uma função que descreve a frequência com que cada nível de cinza aparece na imagem. Ela é utilizada para ver a distribuição de nível de cinza, permitindo que a imagem seja normalizada, excluindo algumas regiões que não são as regiões de interesse pois estão fora da faixa de cinza desejada, eliminando ruídos e principalmente deixando que o sistema fique menos vulnerável a diferentes configurações de scanners que podem gerar imagens com diferentes ajustes de contraste e brilho.

Na figura 4.1, pode ser visto o histograma da imagem original e o histograma da imagem normalizada.

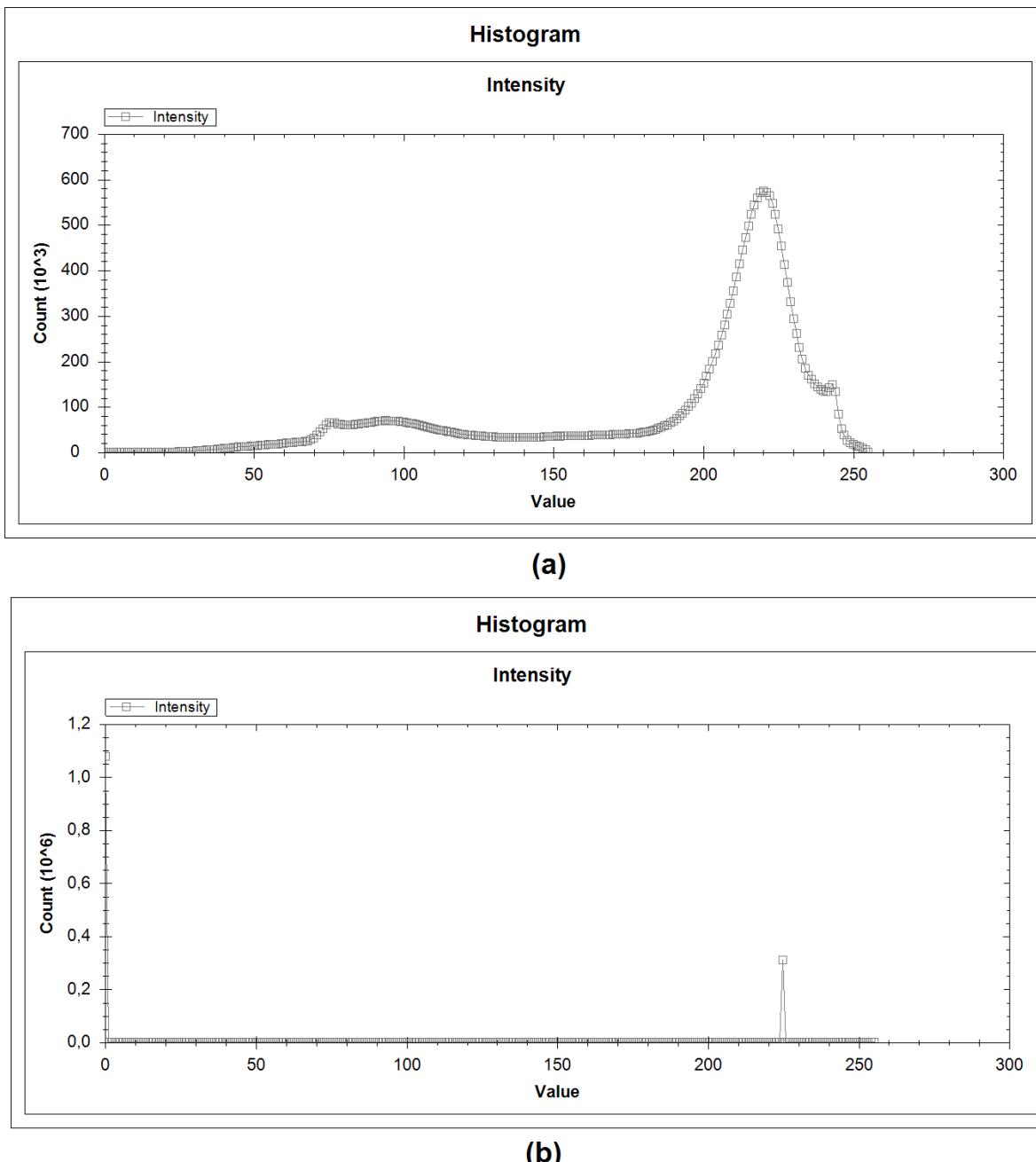


Figura 4.1: a) Histograma da imagem do documento após a conversão para escala de cinza e b) Histograma após a seleção da faixa cinza de interesse

Como pode ser visto, antes da binarização da imagem existia um pequeno pico no valor próximo ao 90 e uma irregularidade após o 240, porém o grande pico era próximo ao 220, após a normalização do histograma pode-se perceber que os pontos fora da faixa de interesse foram ignorados, reduzindo assim o nível de ruído da imagem.

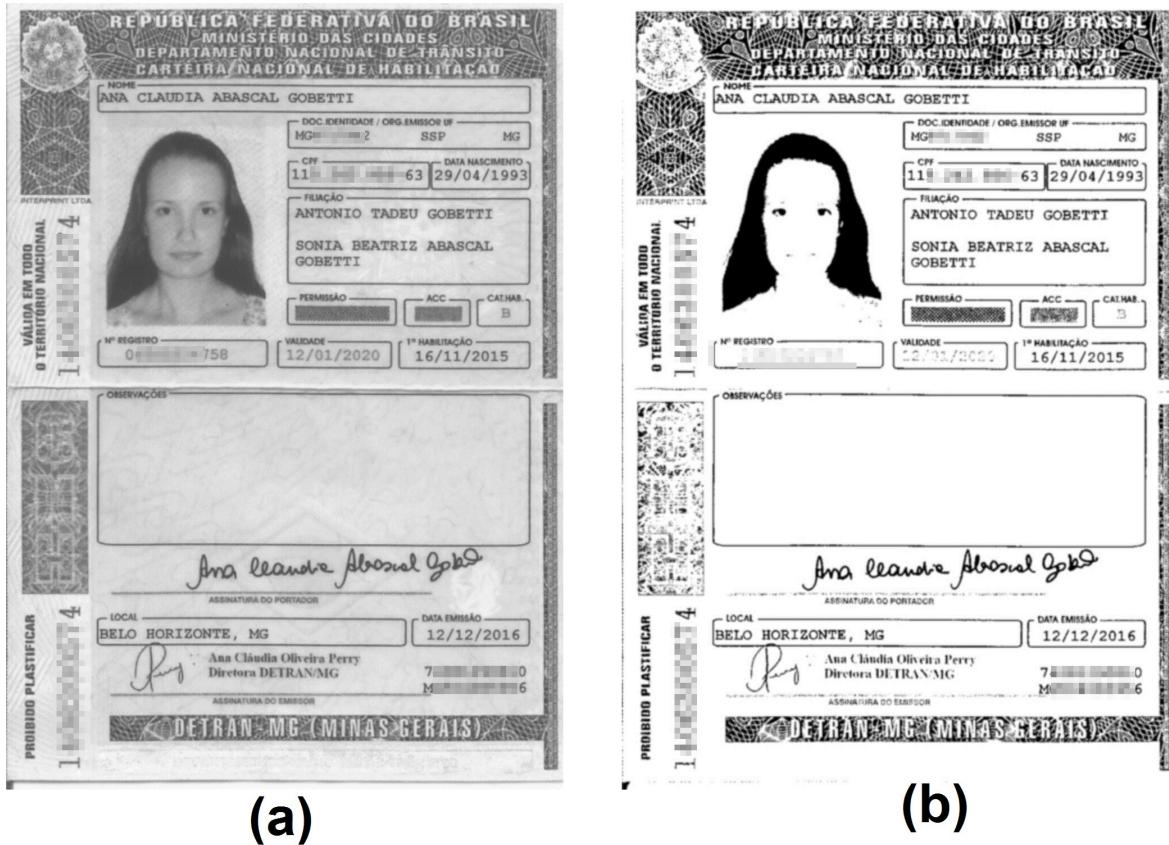


Figura 4.2: a) Imagem da CNH convertida em escala de cinza, sem tratamento e b) após a seleção da faixa cinza de interesse

### 4.1.2 Reconhecimento da face e linhas

Como foi explicado no capítulo anterior, buscou-se identificar os campos de dados com base na localização das linhas e da face contida na CNH, abaixo segue a figura com as delimitações da linha em azul e a delimitação da face em vermelho.

Após este tratamento, foi possível identificar as regiões de interesse e cada uma destas imagens delimitadas pelos retângulos vermelhos da figura 4.3 passou pelo sistema de leitura de caracteres isoladamente e o resultado foi a obtenção dos dados da CNH, estes dados foram enviados para uma interface que será apresentada na próxima seção.



Figura 4.3: Resultado do processamento após a obtenção das linhas

Como um todo, o sistema se mostrou bastante eficiente e após pequenas modificações, como delimitação do conjunto de caracteres do leitor para cada campo, o sistema leu corretamente todas as CNHs testadas. Foram utilizadas 5 (cinco) CNHs do modelo antigo e 1 (uma) CNH do novo padrão. Em média o sistema demorou 7 segundos para exibir os dados na tela e este resultado é muito positivo, uma vez que o tempo para uma pessoa interpretar e digitar o documento é muito superior.

## 4.2 Protótipo da interface

Com o objetivo de deixar o sistema mais flexível a mudanças e atualizações, ele foi dividido em dois módulos completamente independentes. O primeiro modulo é o de identificação e interpretação dos dados, como foi descrito na seção anterior e o segundo modulo é um protótipo do que seria a interface gráfica de utilização do sistema, para este modelo foi realizada uma interface simples somente para seleção da imagem, apresentação dos dados obtidos e arquivamento do mesmo em formato XML.

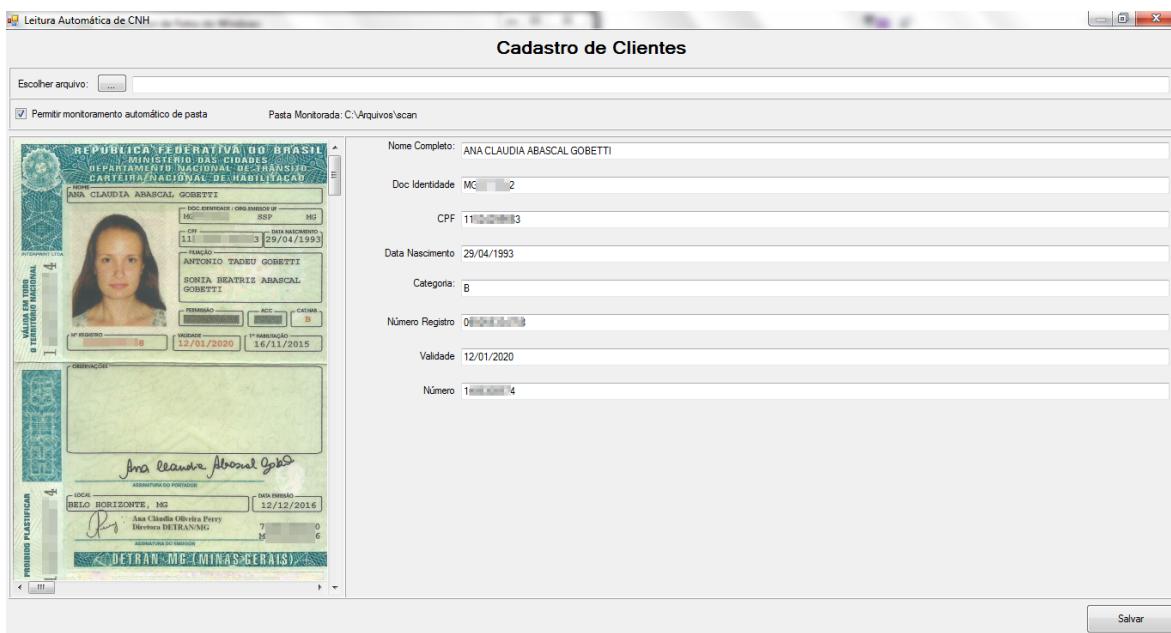


Figura 4.4: Interface desenvolvida para o protótipo

Como foi dito no capítulo anterior, esta interface seria responsável pela seleção da imagem, para este protótipo foram escolhidas duas formas de entrada. A primeira é a seleção manual da imagem, onde o operador do sistema deve navegar até a pasta desejada e escolher a imagem. Além disto, foi implementado no sistema uma função para observar automaticamente uma pasta, que no caso é a pasta de configuração do scanner, esta pasta pode ser alterada nas configurações do sistema e ajuda a tornar o sistema ainda mais rápido, uma vez que a interferência do usuário é menor. Como monitorar uma pasta pode ser uma função não desejada, é possível desabilitar esta opção desmarcando o checkbox na interface.

Uma vez obtida a imagem, é solicitado o tratamento e a interface fica esperando o resultado. Quando recebe o mesmo ela é exibida preenchendo o campo da imagem e os respectivos campos de texto contidos na tela. Estes campos são editáveis, para caso seja necessário realizar alguma correção. Após a conferencia dos dados é possível salvar os arquivos em formato XML, em que o nome do arquivo será formado pelo primeiro nome da pessoa mais o CPF, apertando o botão Salvar no canto inferior direito da tela.

Caso ocorra algum erro no tratamento da imagem, por exemplo, a face não seja identificada, invalidando o documento, o arquivo não seja uma imagem ou a imagem não possua resolução suficiente para o processamento, será exibida um alerta no canto inferior esquerdo.

### 4.3 Leitura de diferentes modelos de CNH

Como foi explicado na seção 2.1, atualmente existem dois modelos de CNH e embora a posição dos campos obedeçam a mesma ordem, a tonalidade dos dois documentos é muito diferente. Apesar disto, o tratamento inicial, em que é realizado o distribuição de cores para filtrar a imagem se mostrou bastante eficiente, sendo possível identificar corretamente os dados dos dois modelos de CNH, como pode ser visto na figura 4.5.

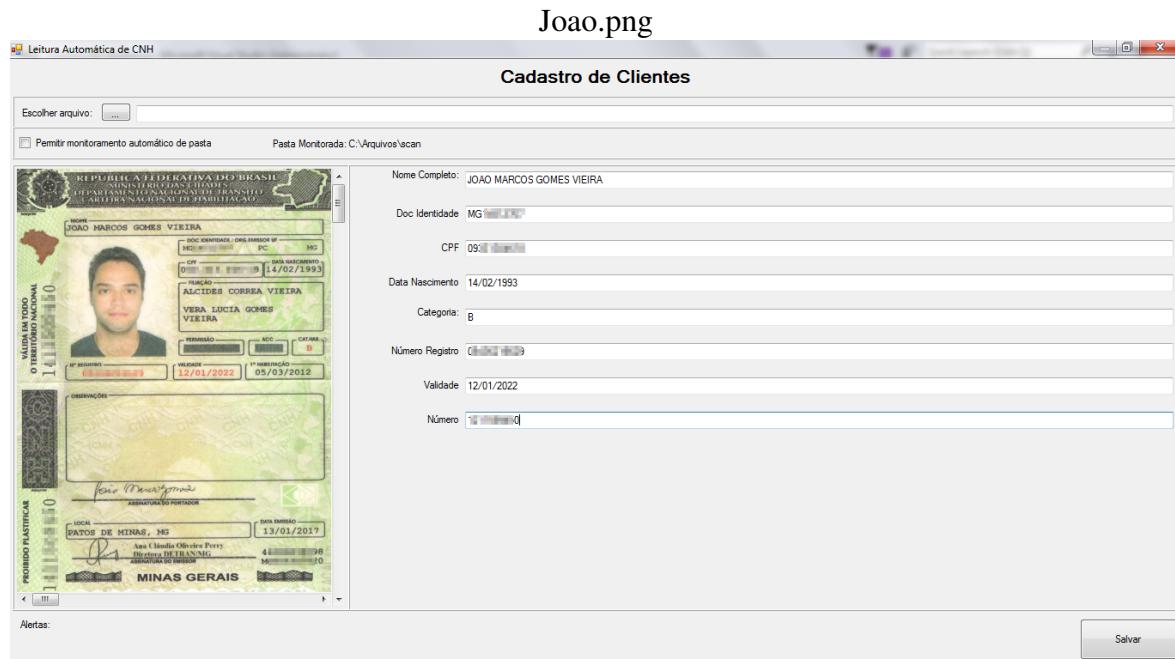


Figura 4.5: Resultados obtidos para o novo modelo de CNH

# Capítulo 5

## Conclusões

### 5.1 Considerações Finais

O objetivo principal deste projeto é a solução de um problema que está despertando a cada dia mais interesse nas empresas, que é a necessidade de se possuir sistemas que agilizem o processo por meio da automação de tarefas, reduzindo não somente o tempo gasto para as mesmas, mas o custo empregado. Atualmente, os sistemas de OCR tradicionais permitem extrair um texto de uma imagem, porém não é atribuído sentido à esta informação extraída. Como discutido a seção 2.4, os sistemas atuais possuem a limitação de estarem presos à uma tecnologia de digitalização de imagens e uma interface gráfica.

Portanto, neste projeto foi proposta a criação de uma ferramenta que além de trazer todos os benefícios da leitura automática ainda possui o diferencial de ser um sistema completamente desacoplado da interface gráfica, possibilitando uma rápida atualização das interfaces, ou mesmo do sistema de reconhecimento, sem que o funcionamento do mesmo seja comprometido.

A estrutura de reconhecimento das regiões de interesse do documento foram definidas depois de pesquisas e testes. A arquitetura definida e implementada mostrou-se bastante eficiente, cumprindo com precisão e rapidez seus objetivos.

O resultado do trabalho é um sistema que está muito próximo de um sistema que pode ser comercializado e aplicado em situações reais para o cadastramento de clientes, como hotéis, locadoras de veículos, entre outros.

### 5.2 Propostas de Continuidade

Com o objetivo de transformar o sistema o mais genérico possível algumas sugestões de continuidade do trabalho:

- Otimização no processamento: O método de processamento da imagem até a identificação de linhas é o mais lento do sistema, comprometendo o desempenho do sistema. A maneira que foi desenvolvida atende às especificações, uma vez que a velocidade de leitura do documento é muito superior ao processo manual, porém este é um ponto em que pode ser realizado uma melhoria;

- Acrescentar um identificador de documentos: este sistema foi desenvolvido com o objetivo de digitalizar a carteira de habilitação, porém existem diversos outros documentos, como documento de identidade, passaporte, carteira de trabalho. Pode-se acrescentar um módulo para reconhecer o documento automaticamente, a partir de modelos já criados;
- Acrescentar um modulo para cadastrar novos modelos: além de tornar o sistema mais genérico, permitindo a identificação e leitura de diversos tipos de documentos, seria interessante criar uma interface para que o próprio usuário dos sistema pudesse cadastrar novos documentos;
- Alterar o método de leitura: Apesar do Tesseract OCR apresentar um bom desempenho, foi observado que antes da criação dos modelos de leitura, ocorriam diversos erros, por exemplo, reconhecia uma letra no campo CPF. Para evitar estes erros, criado três leitores delimitando o conjunto de caracteres possíveis em cada campo. Porém, seria interessante avaliar as outras tecnologias de OCR existentes, com o objetivo de melhorar a qualidade e assertividade do sistema;
- Criar um mecanismo para identificação de fraudes: Atualmente, o número de falsificações de documentos, principalmente a carteira de habilitação é crescente. Isto se deve ao fato que este documento pode ser utilizado como documento de identificação, que permite que a pessoa se passe por outra ou mesmo aumente sua idade. Além disto, é um documento necessário para certificar que a pessoa está habilitado para conduzir aquele veiculo e muitas vezes a pessoa não tem dinheiro para pagar o processo para retirar a documentação de maneira legal e acaba optando pela falsificação. Apesar dos crescentes esforços para evitar as falsificações, que resultam em diversas marcas de segurança nos documentos, as falsificações estão cada vez mais fidedignas. Dessa forma, um sistema que auxiliasse na análise de falsificação dos documentos, conferindo fonte, tamanho da escrita, alinhamento dos campos entre outros fatores, seria de grande utilidade, aumentando a segurança contra pessoas mal intencionadas, como estelionatários;

Dessa forma, este trabalho termina com a sugestão de que este sistema consiga ser uma ferramenta completamente autônoma e genérica, que extraia a informação dos mais variados tipos de documentos, permitindo a criação de novos modelos sem que seja necessário intervenção de um desenvolvedor, facilitando assim, a vida do operador do sistema.

# Referências Bibliográficas

Trustid. URL <https://www.trustid.com/>.

Mitek. URL <https://www.miteksystems.com/>.

EmguCV, 2017. URL [http://www.emgu.com/wiki/index.php/Main{\\_}Page](http://www.emgu.com/wiki/index.php/Main{_}Page).

Aline Gabriel De Almeida, Esther Luna Colombini, Silva Simões, and Campus De Sorocaba. Identificação da posição de objetos circulares coloridos a partir das imagens de uma webcam usando a biblioteca OpenCV Material e Métodos Agradecimentos Resultados e Discussão. page 26.

Alex Araújo. Sistema de hardware e software para verificação de preço em produtos utilizando código de barras. 2015.

Padrões Baseados and E M Imagens. PADRÕES BASEADOS EM IMAGENS DIGITAIS . Fátima Aparecida Tagliaferro Jaguariúna. 2007.

Sven Bertel, Thomas Barkowsky, Dominik Engel, and Christian Freksa. Computational Modeling of Reasoning with Mental Images: Basic Requirements \*. *Review Literature And Arts Of The Americas*, (Iccm):50–55, 2006.

S Bhaskar, N Lavassar, and S. Green. Implementing Optical Character Recognition on the Android Operating System for Business Cards. 2010.

G. BORST and S. M. KOSSLYN. Visual mental imagery and visual perception: Structural equivalence revealed by scanning processes. *Memory & Cognition*, 36(4):849–862, 2008. ISSN 0090-502X. doi: 10.3758/MC.36.4.849. URL <http://www.springerlink.com/index/10.3758/MC.36.4.849>.

G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 2008. ISBN 9780596554040. URL <https://books.google.com.br/books?id=seAgiOfu2EIC>.

FABIO AUGUSTO CABRAL and VICTOR HUGO PEREIRA MACHADO. Identificador de placa veicular: alternativa para segurança em escolas. *Universidade Tecnológica Federal Do Paraná*, 2014.

Amanda Canestraro. Sistema Biométrico Baseado na Análise da Geometria da Mão. page 72, 2013.

- Henrique Carlos and Fonte Boa. Utilização de técnicas de Reconhecimento de Padrões para identificação de ataques de DNS Spoofing em redes locais. 2016.
- Mohamed Cheriet. Optical character recognition devices. 2007.
- CONTRAN. Conselho Nacional de Trânsito. URL <http://www.denatran.gov.br/index.php/contran>.
- Marcos Vinicius da Cruz Correa. Controle de robôs móveis utilizando Kinect. 2014.
- CTB. Código de Trânsito Brasileiro. 1997. URL [http://www.planalto.gov.br/ccivil\\_03/leis/L9503.htm](http://www.planalto.gov.br/ccivil_03/leis/L9503.htm).
- Leonardo Augusto de Oliveira. Localização e reconhecimento de caracteres em placas de automóveis. 2010.
- DETRAN MG. Departamento de Transito de Minas Gerais. 2017. URL <http://www.detran.mg.gov.br>.
- DETRAN PR. Departamento de Transito do Parana. 2017. URL <http://www.detran.pr.gov.br>.
- Jean Dias do Nascimento. Detecção e reconhecimento de placa automotiva com baixo custo. 2012.
- Visuais Em and Imagens Coloridas. Identificação de espécies de peixes utilizando histogramas de palavras visuais em imagens coloridas.
- Fady Aboujaoude. Desenvolvimento de um Sistema de Baixo Custo para Rejeição de Objetos Indesejáveis em Linhas de Produção Usando Visão Computacional. page 71, 2014.
- Edgar Tamio Hirama. Desenvolvimento de aplicação android para reconhecimento óptico de caracteres em brinco de identificação animal. 2014.
- ISO/IEC. Systems and software engineering Life cycle processes Requirements engineering . 2011.
- ILSON AKITO SHIGEHARU JUNIOR. Estudo da viabilidade do uso de um OCR na placa Beagleboard e sua integração no xLupa embarcado. *Universidade Estadual do Oeste do Paraná*, (45), 2014.
- Luiz Fernando Miranda Vieira Lins. Reconhecimento óptico de caracteres (OCR) e análise de sistemas OCR baseados em código aberto. 2012.
- Ludmila de Oliveira Moreira Lopes. Detecção de Movimentos Suspeitos Utilizando Visão Computacional. 2015.

- André Wilson Machado and Bernardo Quiroga Souki. Simplificando a obtenção e a utilização de imagens digitais: scanners e câmeras digitais. *Revista Dental Press de Ortodontia e Ortopedia Facial*, pages 133–156, 2004. doi: 10.1590/S1415-54192004000400012. URL [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1415-54192004000400012&lang=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-54192004000400012&lang=pt).
- Filipe Augusto Manzi. Aplicação de visão computacional para extração de características em imagens do olho humano. page 47, 2007.
- Hugo Mari, José Carlos, and José Carlos Cavalheiro da Silveira. Sobre a cognição visual. *Scripta*, 14(26):3–26, 2010.
- Rui Miguel. Superfícies Interativas com Kinect. 2013.
- S Mori, C.Y. Suen, and K. Yamamoto. Historical review of OCR research and development. 1992.
- Victor Nascimento. Implementação de um sistema de identificação facial utilizando Linux Embarcado. 2015.
- Beatriz Pinheiro and Renato Martins. Carteira de habilitação falsa - Ciência da falsidade ideológica - Venda - Intermediação - Art . 297 do Código Penal - Condenação - Estelionato - Desclassificação - Impossibilidade - Indução a erro do comprador - Não ocorrência - Pena - Dosimetria - Maus a. pages 321–323, 2013.
- Serasa. Serasa Consumidor. URL <https://www.serasaconsumidor.com.br>.
- L. G. SHAPIRO. Computer Vision. 2001.
- Shi Shin. *Emgu CV Essentials*. 2013. URL [http://www.amazon.com/Emgu-CV-Essentials-Shin/dp/1783559527/ref=sr\\_1\\_1?s=books&ie=UTF8&qid=1415059990&sr=1-1&keywords=emgu+cv](http://www.amazon.com/Emgu-CV-Essentials-Shin/dp/1783559527/ref=sr_1_1?s=books&ie=UTF8&qid=1415059990&sr=1-1&keywords=emgu+cv).
- DJ Waters and JLP Júnior. *Do microfilme à imagem digital*. 1997. ISBN 8570090536. URL [http://www.arqsp.org.br/cpba/pdf/\\_cadtec/49.pdf](http://www.arqsp.org.br/cpba/pdf/_cadtec/49.pdf).
- N ZUECH. Understanding and Applying Machine Vision . *Yardley, Pennsylvania: Marcel Dekker INC*, 2 ed, 1988.