

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

Título da Monografia

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

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 possibilitou o aumento da velocidade de processamento dos computadores possibilitando a realização de diversas tarefas que não eram possíveis até pouco tempo, 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 entre informação em formato físico e digital dá origem, a que muitas vezes, seja necessário transferir informações do meio físico para o meio digital, como no caso de alguns sistemas pode ser necessário cadastrar um cliente, digitando o seu nome, número do documento, data

de nascimento, local de nascimento, entre outras informações.

Porém não existe atualmente uma maneira rápida e automática de realizar esta conversão dos dados, sendo necessário um processo 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 processo, reduzindo o tempo de espera em diversas atividades, como em nos processos 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 uma software de visão computacional, com o auxílio da biblioteca OpenCV, capaz de extrair informações de um documento de identificação.

1.3 Organização do Trabalho

Este trabalho está organizado da seguinte forma:

Introdução: Apresenta os sistemas de visão computacional, bem como suas aplicações, vantagens, aplicações e limitações tecnológicas envolvidas. Além disto também são apresentados os objetivos e a organização do trabalho.

Revisão Bibliográfica: Neste capítulo será apresentado 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 descritas todos os passos necessários para implementar o sistema de identificação de fraudes nas CNHs. E serão definidas as formas de avaliação do sistema.

Resultados: Serão apresentados os resultados do sistema implementado no capítulo anterior.

Discussão: nessa seção os resultados obtidos são comparados criticamente com o estado da arte e os conhecimentos existentes.

Conclusão: Serão apresentadas as 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 utilizados na implementação.

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 veículos que o condutor está habilitado a conduzir, sendo elas (DETRAN PR):

- 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.
- C:** condutor de veículo motorizado usado para transporte de carga, com peso bruto superior a 3.500 quilos (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 unidade acoplada ou rebocada tenha peso bruto de 6 mil quilos 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 e ela deve participar de cursos teóricos preparatórios, médico e psicotécnico. Após a primeira habilitação existem algumas regras para mudança de categoria (DETRAN MG):

- Categoria 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). Além disto a CNH tem data de validade, ou seja, motoristas com menos de 65 anos devem renovar a carteira a cada 5 anos, 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.

De acordo com Marcos Traad é crescente o número de pessoas que falsificam este documento para se passar por outras pessoas, mudar identidade ou mudar dados pessoais como data de nascimento ou mesmo dirigir sem ser devidamente habilitado. Para prevenir fraudes "o Detran investe continuamente em tecnologias, para proporcionar maior segurança aos usuários e parceiros nos processos dentro e fora da intuição", como afirma o diretor geral do órgão, Marcos Traad. Dessa forma, a CNH possui diversos mecanismos e marcas de segurança para evitar qualquer tipo de fraude, como pode ser visto na figura X.

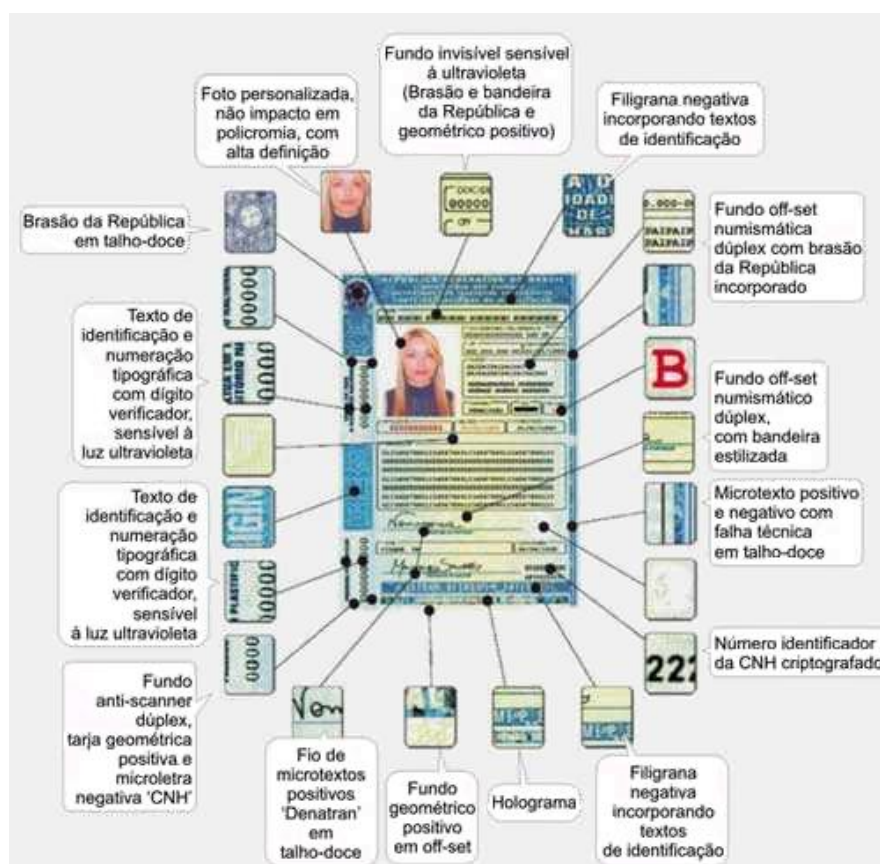


Figura 2.1: CNH Caption

Além disto, a legislação não permite a plastificação do documento, uma vez que como pode ser observado na análise da imagem [] existem pontos na CNH "em talho doce", ou seja, impressão em alto relevo que permitem a identificação de fraudes sem uma perícia minuciosa.

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 da imagem digital permite a codificação digital documentos analógicos, em forma de imagem digital para armazenagem, transmissão em sistemas computadorizados.

Uma imagem digital é a representação de uma imagem bidimensional como uma sequência finita de valores digitais. Cada um destes valores representa um pixel, que é o menor elemento que compõe uma imagem digital ou uma componente de cor de um pixel, no caso de imagens coloridas.

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

- **Imagens Binárias:** Também conhecida como imagens preto e branco. Nestas imagens cada pixel pode assumir valores 0 ou 1. Este tipo de imagem 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).
- **Imagens em escala de cinza:** Nestas imagens 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 coloridas:** Nestas imagens, cada pixel é representado por pelo menos três informações, uma para cada canal de cor, como por exemplo o RGB que será um dos itens discutidos na próxima sessão. O valor de cada canal e composição da imagem colorida depende do espaço de cores utilizado.



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 processamento é a obtenção de informações numéricas ou simbólicas de imagem, como 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 e normalmente são executadas nesta ordem:

1. Aquisição de imagens
2. Pré-processamento: para poder processar as imagens computacionalmente elas precisam ser pré-processadas. Uma imagem digital pode possuir muitas informações que não devem ser utilizadas e podem influenciar negativamente na tomada de decisão do sistema. O pré-processamento permite que estas informações indesejadas sejam ignoradas ou retiradas (ZUECH, 1988).
3. Extração de características: Extração de objetos de interesse e partes da imagem.
4. Processamento de alto nível: Nesta etapa a entrada é tipicamente um pequeno conjunto de dados, como um conjunto de pontos ou uma região da imagem que se supõe conter um objeto específico (DAVIES, 2005). Este processamento permite estimar parâmetros específicos da aplicação, como orientação e tamanho do objeto de interesse, classificação dos objetos, etc.
5. Tomada de decisão

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 chamada RGB (Red, Green, Blue) e ela produz várias cores através das três cores básicas: vermelha, verde e azul. Nesta 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 colocar 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 do mesmo.

2.3.2 Threshold

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

O threshold é um valor limite que é usado 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 e caso sejam maiores ou iguais a esse limiar ele é transformado em pretos (valor 0) e transformando em brancos (valor 255), caso contrário.

2.3.3 Smooth

A operação de smooth busca suavizar uma imagem eliminando pontos que estejam fora de um padrão esperado. A suavização consiste na aplicação de um filtro passa-baixa para retirada de ruídos. O filtro de média possui uma janela de determinado tamanho que irá percorrer toda a imagem e o elemento central dessa janela receberá a média de todos os elementos da janela. Caso o valor do pixel seja muito diferente do valor médio dos pixels vizinhos, ele é interpretado como ruído e seu valor é corrigido.

2.4 Sistemas OCR

Segundo (Bhaskar, 2010), o algoritmo de reconhecimento óptico 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 e finalmente reconhecimento de caligrafias manuais.

Segundo (Mori, 1992) os primeiros avanços destas tecnologias começaram nos anos 50, inicialmente apenas conceitualmente, mas gradualmente avançando para ferramentas mais concretas. As primeiras versões eram bastante sujeitas à falhas e se baseavam em comparação 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 identificação dos mesmo. Esta mudança de abordagem deu origem aos sistemas atuais.

Capítulo 3

Materiais e Métodos

Neste capítulo, você deve apresentar uma breve revisão bibliográfica sobre as técnicas utilizadas para solução do problema.

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 (AABY, 2004).

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 (OPENCV, 2013). 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 sido 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 (Rui Miguel, 2013 - Superfícies Interativas com Kinect)

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, 2012).

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.