



Instituto Superior de Engenharia de Lisboa
Área Departamental de Engenharia de Electrónica e
Telecomunicações e de Computadores (ADEETC)

Sistema de Digitalização, Detecção e Classificação Automática de Recibos em Dispositivos Móveis

MIGUEL JOÃO RODRIGUES BATISTA
Engenheiro

Trabalho Final de Mestrado para obtenção do grau de Mestre em Engenharia
de Redes de Comunicação e Multimédia

Orientadores:

Doutor Pedro Miguel Torres Mendes Jorge
Doutor Rui Manuel Feliciano de Jesus
Engenheiro Pedro Miguel Marques Morais

Júri:

Presidente: Doutor Paulo Manuel Trigo Cândido da Silva

Vogais:

Doutor Nuno Manuel Robalo Correia
Doutor Pedro Miguel Torres Mendes Jorge
Doutor Rui Manuel Feliciano de Jesus

Novembro de 2012



Instituto Superior de Engenharia de Lisboa
Área Departamental de Engenharia de Electrónica e
Telecomunicações e de Computadores (ADEETC)

Sistema de Digitalização, Detecção e Classificação Automática de Recibos em Dispositivos Móveis

MIGUEL JOÃO RODRIGUES BATISTA
Engenheiro

Trabalho Final de Mestrado para obtenção do grau de Mestre em Engenharia
de Redes de Comunicação e Multimédia

Orientadores:

Doutor Pedro Miguel Torres Mendes Jorge
Doutor Rui Manuel Feliciano de Jesus
Engenheiro Pedro Miguel Marques Morais

Júri:

Presidente: Doutor Paulo Manuel Trigo Cândido da Silva

Vogais:

Doutor Nuno Manuel Robalo Correia
Doutor Pedro Miguel Torres Mendes Jorge
Doutor Rui Manuel Feliciano de Jesus

Novembro de 2012

Resumo

Ao longo dos últimos anos tem-se assistido a uma rápida evolução dos dispositivos móveis. Essa evolução tem sido mais intensa no poder de processamento, na resolução e qualidade das câmaras e na largura de banda das redes móveis de nova geração. Outro aspecto importante é o preço, pois cada vez mais existem dispositivos móveis avançados a um preço acessível, o que facilita a adopção destes equipamentos por parte dos utilizadores. Estes factores contribuem para que o número de utilizadores com “computadores de bolso” tenda a aumentar, possibilitando cada vez mais a criação de ferramentas com maior complexidade que tirem partido das características desses equipamentos.

Existem muitas aplicações que exploram estas características para facilitar o trabalho aos utilizadores. Algumas dessas aplicações conseguem retirar informação do mundo físico e fazer algum tipo de processamento, como por exemplo, um leitor de códigos QR ou um OCR (Optical Character Recognizer).

Aproveitando o potencial dos dispositivos móveis actuais, este trabalho descreve o estudo, a implementação e a avaliação de uma aplicação de realidade aumentada para adquirir e gerir recibos em papel de forma automática e inteligente. A aplicação utiliza a câmara do dispositivo para adquirir imagens dos recibos de forma a poder processá-las recorrendo a técnicas de processamento de imagem. Tendo uma imagem processada do recibo é efectuado um reconhecimento óptico de caracteres para extracção de informação e é utilizada uma técnica de classificação para atribuir uma classe ao documento. Para um melhor desempenho do classificador é utilizada uma estratégia de aprendizagem incremental. Após a correcta classificação é possível visualizar o recibo com informação adicional (realidade aumentada). O trabalho proposto inclui também a avaliação da interface e dos algoritmos desenvolvidos.

Palavras-chave: Aplicações em Dispositivos Móveis, Processamento de imagem, Reconhecimento óptico de caracteres, Realidade Aumentada

Abstract

Over the last few years mobile devices have evolved quickly. That evolution has been occurring in the fields of processing power, camera image quality and resolution, and in new generation high bandwidth wireless networks. Cost is another important aspect, because the number of advanced mobile devices at affordable prices has been growing, making it easier for users to adopt these devices. These factors contribute to the increase in the number of users with “handheld computers”, paving the way for new and more complex tools that harness the new capabilities of those devices.

There are many apps (mobile device applications) that use those capabilities to ease the user’s work. Some of those apps are able to retrieve information from the physical world and process it, for example, like a QR code reader or an OCR (Optical Character Recognizer).

Harnessing the potential of today’s mobile devices, this document describes the study, implementation and evaluation of an augmented reality app that automatically and intelligently processes and classifies paper receipts. The app will use the device’s camera to capture images of the receipts so that it can apply image processing techniques on them. An OCR will be used after the image has been processed to extract the information that it contains and a computer learning technique will be used to classify the document. An incremental learning strategy will be used for a better classifier performance. It is possible to visualize the receipt with additional information (augmented reality) after its correct classification. The proposed work also includes the evaluation of the user interface and the developed algorithms.

Key words: Mobile Device Apps, Image Processing, Optical Character Recognition, Augmented Reality

Índice

Introdução.....	1
1.1 Motivação e objectivos.....	2
1.2 Enquadramento.....	4
1.3 Solução apresentada	9
1.4 Contribuições.....	13
1.5 Organização da Dissertação.....	13
Trabalho Relacionado.....	15
2.1 Gestão de Recibos em Dispositivos Móveis	15
2.2 Reconhedor Óptico de Caracteres	17
2.2 Processamento de Recibos no Servidor versus Cliente	19
2.3 Processamento de Imagem de Documentos	19
2.4 Realidade Aumentada.....	22
2.5 Tecnologias.....	25
Digitalização, Detecção e Classificação Automática de Recibos	28
3.1 Diagrama de Blocos	28
3.2 Aquisição e Detecção de Recibos	29
3.3 Extracção de Informação.....	33
3.3.1 Reconhecimento Óptico de Caracteres	34
3.3.2 Extracção de Valores.....	34
3.4 Classificação.....	37
Resultados.....	40
4.1 Características dos Dados	40
4.2 Ambiente de Testes.....	40
4.3 Tempos de Execução	41
4.4 OCR	43
4.5 Transformação Homográfica.....	48
4.6 Classificador	49
4.7 Interface Gráfica.....	52

Conclusões e Trabalho Futuro	55
Bibliografia	58
Apêndice	58
A.1 Diagrama de Classes	58
A.2 Interface Gráfica e Utilização	58

Lista de Figuras

Figura 1 – Processo de reconhecimento e captura de um recibo.	3
Figura 2 – Distribuição de <i>smartphones</i> no mercado Norte-americano no 3º trimestre de 2011.	5
Figura 3 – Quotas de mercado dos principais fabricantes de <i>smartphones</i>	5
Figura 4 – Evolução da quota de mercado dos <i>smartphones</i> nos EUA de Outubro de 2010 a Fevereiro de 2012.	6
Figura 5 – Distribuição dos principais sistemas operativos para <i>smartphones</i>	7
Figura 6 – Crescimento mensal dos principais sistemas operativos para dispositivos móveis nos EUA entre Janeiro de 2010 e Junho de 2012.	8
Figura 7 – Diagrama de blocos do sistema.	9
Figura 8 – Shoeboxed Receipt Tracker and Reader.	16
Figura 9 – Aplicação Expensify.	17
Figura 10 – Image to Text – OCR.	18
Figura 11 – FotoNote.	18
Figura 12 – JotNot Scanner.	20
Figura 13 – Genius Scan.	21
Figura 14 – Google Goggles.	23
Figura 15 – Layar.	24
Figura 16 – Word Lens.	25
Figura 17 – Diagrama de Blocos do Sistema Simplificado.	29
Figura 18 – Exemplo de Detecção de Recibo.	30
Figura 19 – Diagrama de blocos da Aquisição e Detecção do Recibo.	31
Figura 20 – Filtragem com filtro de média (à esquerda); Binarização (à direita).	31
Figura 21 – Detecção de contornos de Canny (à esquerda); Detecção dos limites do recibo com base na Transformada de Hough (à direita).	32
Figura 22 – Transformação homográfica (à esquerda); <i>Crop</i> do recibo (à direita).	33
Figura 23 – Diagrama de blocos da Extração de Informação.	33
Figura 24 – Diagrama de blocos do Reconhecimento Óptico de Caracteres.	34
Figura 25 – Diagrama de blocos da Extração de Valores.	34
Figura 26 – Exemplo de distância de Hamming entre termos.	35

Figura 27 – Exemplos de recibos e dados a extrair.....	36
Figura 28 – Diagrama de blocos da Classificação.....	37
Figura 29 – Aquisição do recibo com realidade aumentada.	41
Figura 30 – Resultado da extracção de valores e classificação do recibo com possibilidade de edição pelo utilizador.	42
Figura 31 – Visualização do resultado antes de guardar o recibo no dispositivo móvel.	42
Figura 32 – Exemplo de recibo com margens grandes.....	47
Figura 33 – Matriz de transformação afim do Core Graphics.	48
Figura 34 – Probabilidade de sucesso e insucesso do classificador ao longo das iterações.	51
Figura 35 – Exemplo de realidade aumentada para detecção dos contornos do recibo.	53
Figura 36 – Sobreposição do resultado do processamento do recibo.	54
Figura 37 – Diagrama de classes da aplicação.....	58
Figura 38 – Menu inicial.	59
Figura 39 – Lista de recibos já classificados.....	60
Figura 40 – Detalhes de um recibo já classificado.....	60
Figura 41 – Lista de termos mais prováveis por classe.	61
Figura 42 – Processamento de fotografias antes de escolher uma imagem de um recibo.....	62
Figura 43 – Menu para carregar imagens de recibos ou iniciar modo passo-a-passo.	63
Figura 44 – Modo passo-a-passo.	64
Figura 45 – Processamento e classificação de imagem terminados.	65
Figura 46 – Modo vídeo: detecção de contornos em tempo real com realidade aumentada.....	66
Figura 47 – Modo vídeo: Processamento e classificação do recibo terminados.	67

Lista de Tabelas

Tabela 1 – Duração média de execução de cada processo.	43
Tabela 2 – Duração média do OCR no dispositivo móvel.	44
Tabela 3 – Duração média do OCR no servidor.	45
Tabela 4 – Dimensão média do ficheiro e duração média de transferência.....	45
Tabela 5 – Qualidade do texto detectado pelo OCR em imagens a cores por dimensão horizontal do recibo.....	46
Tabela 6 – Qualidade do texto detectado pelo OCR em imagens em níveis de cinzento por dimensão horizontal do recibo.....	46
Tabela 7 – Qualidade do texto detectado pelo OCR em imagens binárias por dimensão horizontal do recibo.....	46
Tabela 8 – Matriz de Confusão para a 1ª iteração.....	50
Tabela 9 – Matriz de Confusão para a 2ª iteração.....	50
Tabela 10 – Matriz de Confusão para a 3ª iteração.	50
Tabela 11 – Matriz de Confusão para a 4ª iteração.	50
Tabela 12 – Matriz de Confusão para a 5ª iteração.	50
Tabela 13 – Duração média dos processos envolvidos na detecção com realidade aumentada.....	52



Introdução

A rápida evolução das redes móveis, da capacidade de processamento e da qualidade das câmaras dos dispositivos móveis nos últimos anos, assim como a massificação dos *smartphones*¹, faz com que grande parte da população tenha um computador no bolso, o que abre inúmeras oportunidades para novas aplicações que até então pertenciam ao domínio dos computadores de secretária.

Em termos de processamento, temos vindo a assistir a um aumento da capacidade dos *smartphones* e *tablets*² que são colocados no mercado – a Lei de Moore [1] também se está a verificar no campo dos dispositivos móveis com o lançamento de novos processadores a um ritmo rápido. Nos últimos dois anos evoluiu-se de processadores *single-core* para processadores *dual-core* e já foram lançados processadores *quad-core* no final de 2011.

Outro aspecto importante é a evolução das redes móveis, cuja largura de banda tem vindo a aumentar rapidamente possibilitando elevadas taxas de transferência de dados, que permite aplicações desde o envio e recepção de áudio e vídeo em grande qualidade para, por exemplo, vídeo chamadas ou transmissão de canais de televisão, até à utilização da *cloud*³ para armazenamento remoto de ficheiros de grandes dimensões ou uso de outros serviços remotos. A chegada das redes 4G irá alargar ainda mais o leque de possíveis utilizações da rede num dispositivo móvel, assim como possibilitar o

¹ Telefones móveis com capacidades avançadas de comunicação e processamento, e cujas funcionalidades podem ser estendidas através de novas aplicações

² Dispositivo móvel maior que um *smartphone* mas com funcionalidades semelhantes, e cujo principal meio de interacção é através de um ecrã táctil

³ Conjunto de recursos e serviços disponibilizados por um servidor remoto

uso de processamento remoto com mais frequência, uma vez que o tempo de envio de dados, processamento no servidor e recepção dos resultados deverá começar a aproximar-se, ou até mesmo ser inferior, ao tempo de processamento no próprio dispositivo.

Segundo dois estudos da *International Data Corporation* (IDC) [2] e [3], no último trimestre de 2011 os *smartphones* ultrapassaram pela primeira vez os PCs como tipo de dispositivos mais utilizados, com cerca de 100,9 e 92,1 milhões de unidades vendidas respectivamente. Outro estudo por parte da *NetMarketShare* [4], relacionado com acesso à Internet, conclui que de entre todos os sistemas operativos de *smartphones* e *tablets*, o sistema operativo mais utilizado no acesso à Web é o iOS com 54% de taxa de utilização.

O trabalho proposto neste documento tem como objectivo construir uma aplicação que tire partido das actuais potencialidades dos dispositivos móveis para resolver o problema da gestão automática de recibos em papel.

1.1 Motivação e objectivos

Num mundo cada vez mais digital ainda existem alguns processos que usam ou dependem do formato em papel. Os recibos, por exemplo, embora já eles próprios criados digitalmente através de inúmeros programas de facturação, continuam a ser impressos em papel, e assim entregues ao cliente, para depois este ter de o digitalizar de forma a melhor organizar os seus gastos, preencher o seu IRS ou apresentar as suas despesas à empresa onde trabalha, entre outras situações.

Actualmente existem inúmeras aplicações que tiram proveito das características dos novos dispositivos móveis. No campo do processamento de imagem, existem exemplos como leitores de códigos de barras, leitores de *QR Codes*, aplicações que melhoram a qualidade das fotografias tiradas, entre outras. Existem ainda algumas aplicações que extraem informação de recibos em formato de papel [5]. Contudo, o processamento do recibo não é feito no dispositivo e o resultado desse processamento não está disponível de imediato ao utilizador, ou então o processo não é tão automático quanto o desejável e o utilizador tem de introduzir a informação manualmente.

Assim, de forma a aproveitar todo o potencial da mobilidade e poder computacional dos dispositivos móveis actuais, é objectivo desta dissertação desenvolver um sistema que facilite a gestão de recibos de forma automatizada e inteligente, facilitando o trabalho dos cidadãos na gestão das suas contas. Este sistema deverá ser o mais natural e transparente possível para que possa ser utilizado sem dificuldade tanto por utilizadores avançados como por utilizadores que não tenham muito contacto com a tecnologia.

O sistema proposto, de uma forma geral, irá adquirir o recibo através da câmara do dispositivo móvel, sendo o recibo em formato digital processado e classificado utilizando técnicas de processamento de imagem e de classificação. A Figura 1 ilustra o processo que é composto pelas seguintes etapas:

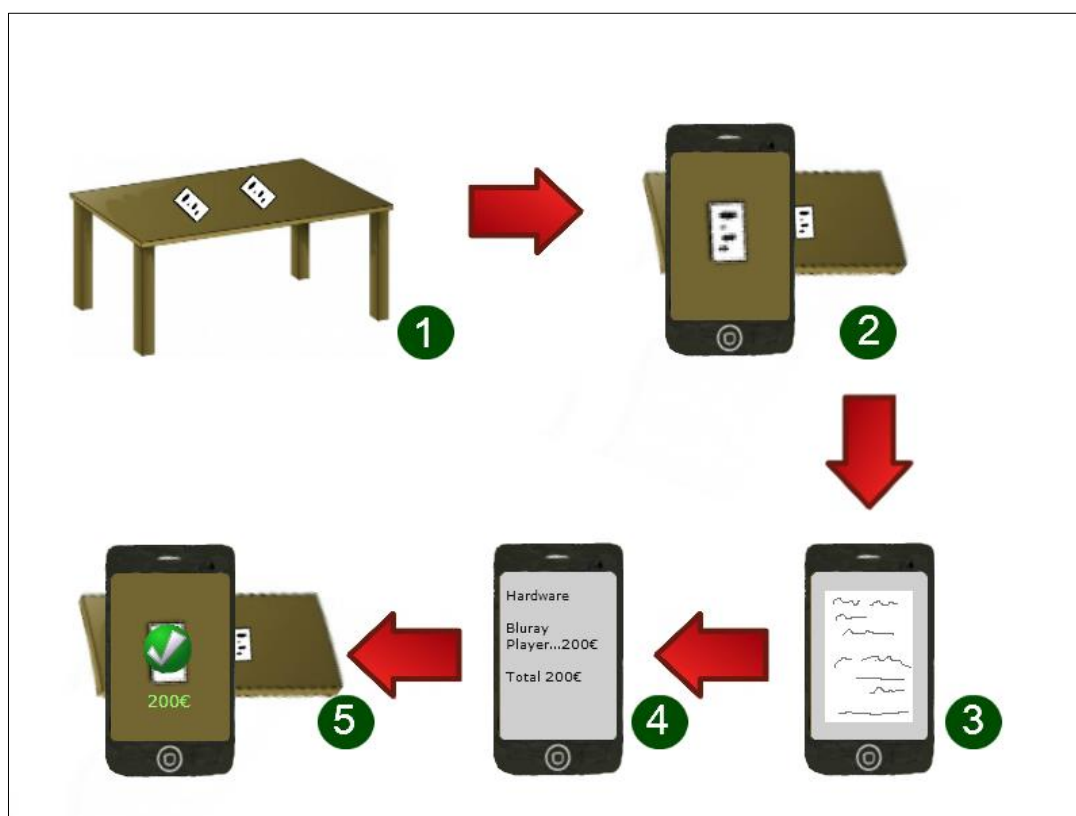


Figura 1 – Processo de reconhecimento e captura de um recibo.

1. O utilizador coloca os recibos sobre uma superfície adequada;
2. Utilizando a aplicação, o utilizador aponta a câmara do dispositivo para um recibo, enquadrando-o;

3. É feita aquisição, segmentação e a normalização da imagem;
4. Processa-se a imagem normalizada com um Reconhecedor Óptico de Caracteres (OCR – *Optical Character Recognizer*), de forma a extrair as informações relevantes contidas no recibo e utiliza-se parte dessa informação para efectuar a classificação do recibo;
5. Utiliza-se realidade aumentada para dar indicação ao utilizador sobre o resultado do processamento.

Por fim, todos os dados extraídos, imagem normalizada e o resultado da classificação são armazenados no dispositivo.

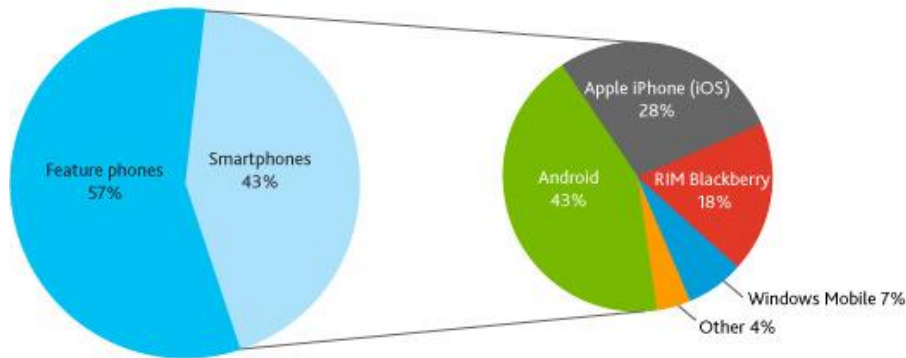
1.2 Enquadramento

Segundo um estudo de mercado [6], os dispositivos móveis são os novos dispositivos de eleição no que diz respeito ao acesso à Internet. Este facto tem-se vindo a verificar ao longo dos últimos anos e apresenta um crescimento acelerado.

Segundo um estudo da Nielsen [7], nos Estados Unidos da América os dispositivos móveis mais populares da Apple, nomeadamente o iPad, iPhone e iPod Touch, estão no topo da lista de preferências para a época Natalícia de 2011. Outro estudo, por parte da Gartner [8], identifica as lojas de aplicações móveis como a *App Store* da Apple ou a *Play Store* da Google como uma das mais importantes tecnologias estratégicas para as empresas em 2012, prevendo que por volta de 2014 o número de *downloads* de aplicações – ou Apps – ronde os 70 mil milhões anuais. A Nielsen publicou ainda outro estudo [9] e [10] que mostra que os dispositivos Android e o iPhone são os líderes no mercado dos *smartphones* no 3º trimestre de 2012 (ver Figura 2), e que de entre de todos os fabricantes, a Apple é o fabricante número 1 com 28% da quota de mercado dos sistemas operativos móveis (ver Figura 3).

Smartphone Penetration and OS Share

Q3 2011, U.S.



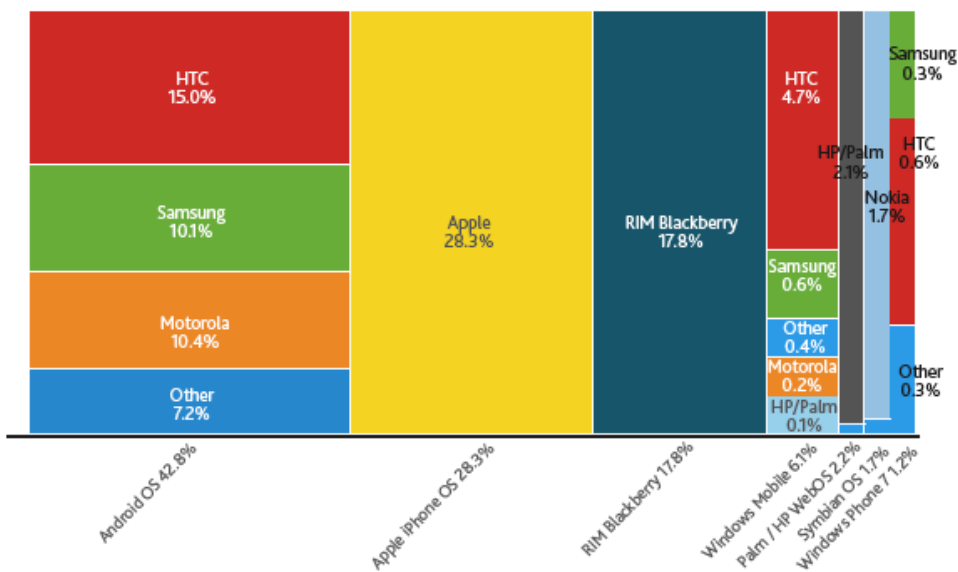
Source: Nielsen

nielsen

Figura 2 – Distribuição de *smartphones* no mercado Norte-americano no 3º trimestre de 2011.

Manufacturer operating system share-smartphones

Q3 '11; postpaid mobile subscribers, n=22,200



Source: Nielsen

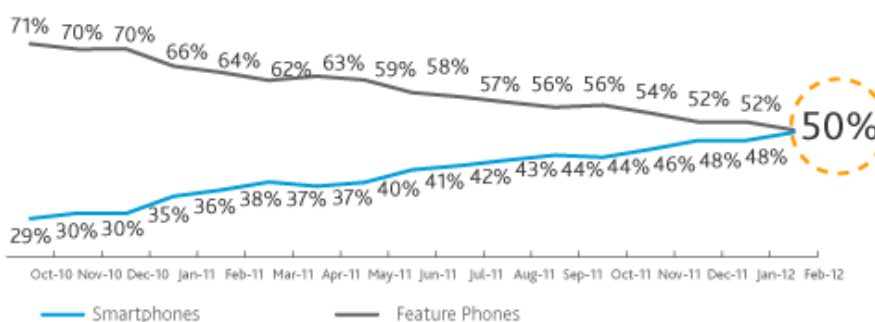
nielsen

Figura 3 – Quotas de mercado dos principais fabricantes de *smartphones*.

Em Março de 2012, a Nielsen publicou os resultados de outro estudo [11] que comprova a tendência registada no estudo anteriormente mencionado, com os *smartphones* a totalizarem quase 50% dos telefones móveis nos Estados Unidos da América em Fevereiro de 2012 (Figura 4).

U.S. Smartphone Penetration

February 2012, Nielsen Mobile Insights



Read as: During February 2012, 50 percent of US mobile subscribers owned a smartphone

Source: Nielsen

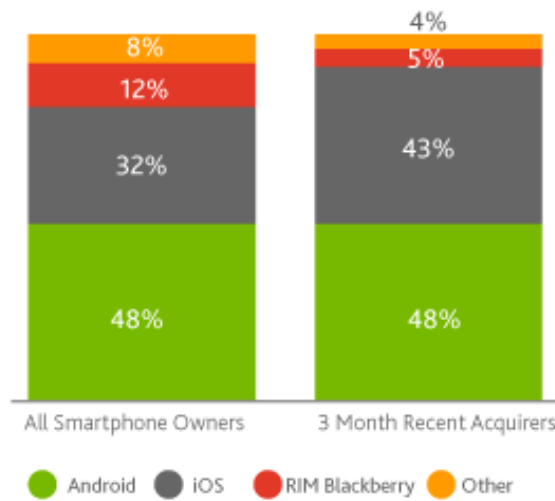
nielsen

Figura 4 – Evolução da quota de mercado dos *smartphones* nos EUA de Outubro de 2010 a Fevereiro de 2012.

O mesmo estudo refere que aproximadamente 90% dos *smartphones* comprados no primeiro trimestre de 2012 são Android ou iOS. Na Figura 5 é possível ver um gráfico que representa essa distribuição, em que o Android e o iOS dominam o mercado. Contudo, também é possível verificar que as pessoas que compraram um novo *smartphone* entre Dezembro de 2011 e Fevereiro de 2012 optaram mais vezes por dispositivos iOS do que Blackberry ou outros, uma vez que a percentagem de iOS aumentou em detrimento do Blackberry e outros.

Smartphone Operating System Share

February 2012, Nielsen Mobile Insights



Read As: During February 2012, 48 percent of smartphone owners had a device that runs on the Android operating system

Source: Nielsen

nielsen

Figura 5 – Distribuição dos principais sistemas operativos para *smartphones*.

Na Figura 6 é apresentado outro gráfico publicado pela Asymco [12] baseado em dados da comScore [13], onde é possível verificar a tendência de crescimento dos vários sistemas operativos para dispositivos móveis nos EUA. Fazendo uma rápida análise ao gráfico da Figura 6 (também visível na Figura 5), conclui-se os sistemas operativos com maior crescimento são o Android e o iOS.

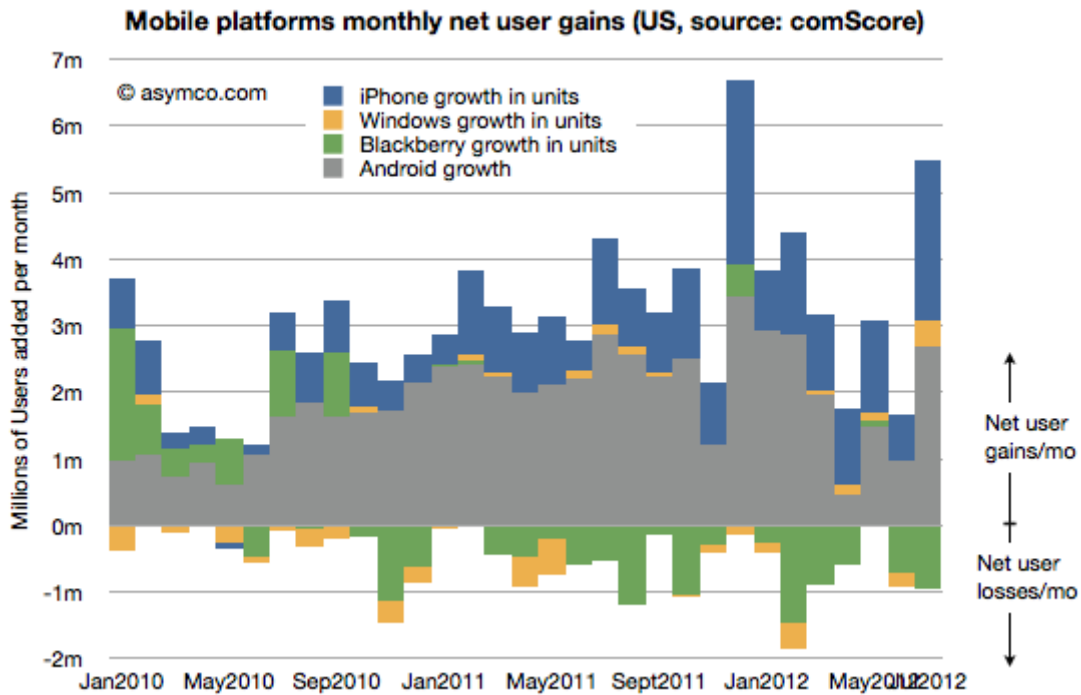


Figura 6 – Crescimento mensal dos principais sistemas operativos para dispositivos móveis nos EUA entre Janeiro de 2010 e Junho de 2012.

Para a elaboração desta dissertação, optou-se por desenvolver o Sistema de Reconhecimento e Gestão de Recibos para dispositivos iOS, nomeadamente iPhone e iPod Touch. As razões desta escolha centram-se nos estudos apresentados até aqui, que dão relevância à plataforma iOS, pelo conhecimento técnico da plataforma por parte dos orientadores, e pelo enquadramento profissional do autor.

1.3 Solução apresentada

O trabalho desenvolvido nesta dissertação teve uma fase inicial de estudo sobre as técnicas de classificação, principalmente sobre os métodos mais utilizados no reconhecimento óptico de caracteres e na classificação de documentos, e sobre várias tecnologias de realidade aumentada para dispositivos móveis. Após esta fase inicial foram definidos os requisitos do sistema, e de seguida foi construída a estrutura da aplicação que após várias iterações resultou na solução apresentada na Figura 7. Esta figura mostra o diagrama de blocos da solução proposta neste documento que permite uma visualização geral das várias partes que compõe o sistema.

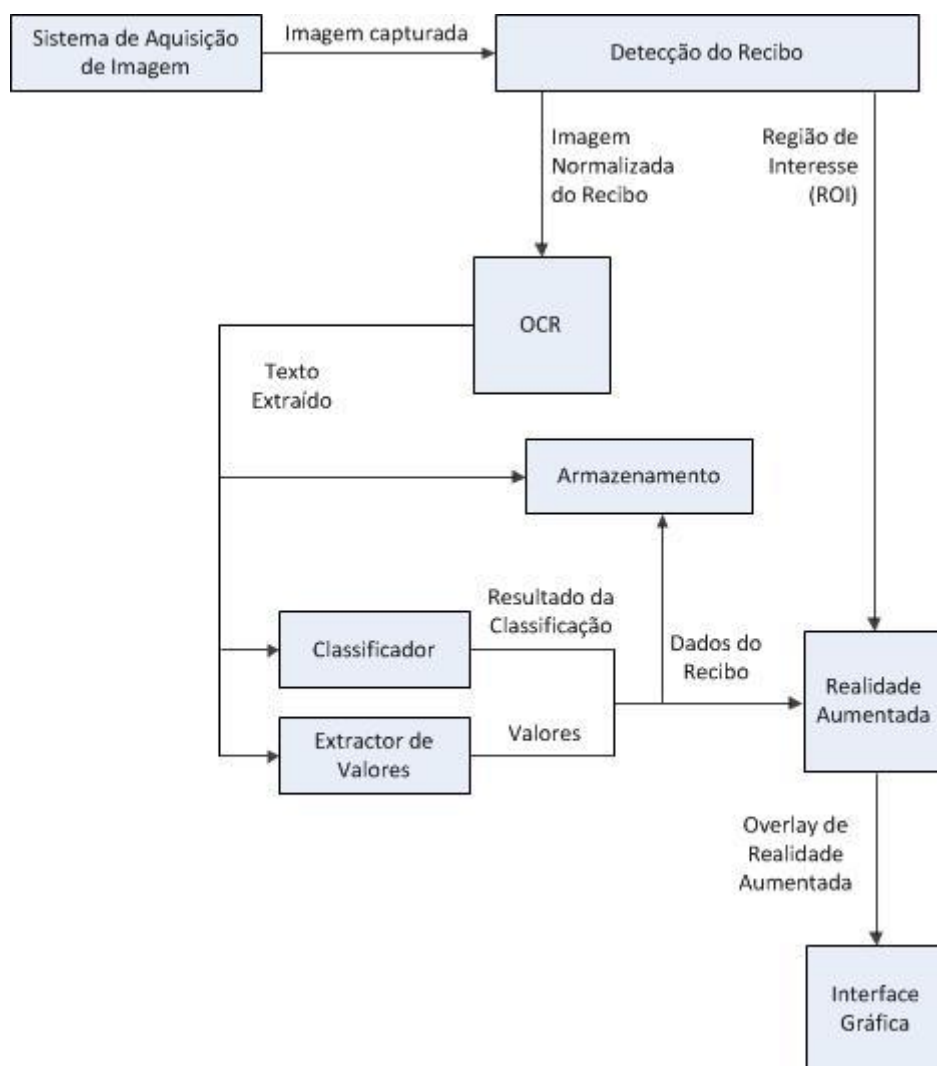


Figura 7 – Diagrama de blocos do sistema.

O sistema desenvolvido é composto pelos seguintes blocos (ver Figura 7):

- **Sistema de Aquisição de Imagem** – composto pela própria câmara do dispositivo móvel, serve para obter uma imagem do recibo.
- **Detecção do Recibo** – bloco responsável por todo o processamento de imagem, ou seja, identificar a posição do recibo na imagem e obter a imagem normalizada do mesmo.
- **Extracção de Informação** – bloco responsável por extrair o texto do recibo e processá-lo de forma a determinar os valores relativos ao total do recibo e ao número de identificação fiscal (NIF) da entidade emissora do recibo.
 - **OCR** – sub-bloco responsável pelo reconhecimento óptico de caracteres de forma a extrair o texto contido no recibo.
 - **Extracção de valores** – sub-bloco responsável por processar o texto devolvido pelo OCR e extrair valores como o NIF da entidade emissora ou o valor total do recibo.
- **Classificador** – este bloco processa o texto extraído pelo OCR e classifica-o numa das classes predefinidas usando técnicas de aprendizagem automática.
- **Gestão de recibos** – bloco responsável pelo armazenamento da informação, está dividido em duas partes:
 - **Armazenamento de Recibos** – armazena a informação extraída dos recibos, como por exemplo total e NIF, o resultado da classificação e a imagem do recibo.
 - **Armazenamento de Palavras e Classes** – armazena o texto dos recibos associados a cada classe e a informação relacionada com a classificação, como por exemplo, probabilidades das classes e probabilidades condicionadas das palavras.
- **Interface Gráfica** – bloco que permite ao utilizador interagir com o sistema, mostrando o resultado das operações, permitindo ver todos os recibos já classificados e obter outras informações.

- **Realidade Aumentada** – sub-bloco responsável por dar informação ao utilizador sobrepondo informação virtual sobre a imagem do mundo real.

Para concretizar os objectivos propostos identificaram-se os seguintes problemas a resolver (Figura 7):

- Segmentação de Imagem – **detectar os recibos a partir de uma imagem ou de vídeo em tempo real**. Para tal é necessário aplicar filtros de contornos de modo a detectar onde se encontram os recibos, assumindo que estes têm contornos rectangulares. Esta detecção deverá facilitar a segmentação da imagem de modo a reduzir a área de imagem a processar.
- OCR – aplicar um **OCR sobre a imagem normalizada dos recibos** e extrair informações relevantes contidas nos mesmos.
- Classificador – utilizar métodos de classificação para **categorizar cada um dos recibos** de modo a conseguir organizar os mesmos da melhor forma possível, atribuindo uma classe a cada recibo, como combustível, restauração, saúde, supermercado, entre outras classes possíveis. Pode ser utilizado, por exemplo, o classificador Naive Bayes com um treino inicial por parte do utilizador, sendo que posteriormente o processo de classificação deverá ser automático devido ao crescimento do conjunto de treino, composto pelos recibos já classificados.
- Realidade Aumentada – para dar indicações ao utilizador em relação à informação extraída, ou mesmo em relação ao sucesso ou insucesso do processo de extracção/reconhecimento, são usadas técnicas de **realidade aumentada para mostrar informação sobre a imagem real**.

A solução proposta inclui também o desenvolvimento de uma interface gráfica para dispositivos móveis.

No final o utilizador tem uma imagem normalizada do recibo e um registo digital da informação nele contidas. Os recibos são armazenados no dispositivo e organizados segundo a categoria em que foram classificados, por exemplo, alimentação ou combustíveis.

Em relação à arquitectura do sistema, existem várias abordagens possíveis neste contexto. Num extremo o dispositivo móvel tem capacidade de processamento suficiente e fica encarregue de todo o processo; noutro extremo o dispositivo móvel serve apenas como dispositivo de aquisição de imagem e apresentação de resultados, enquanto todo o processamento é feito remotamente num servidor.

No entanto, é preciso ter em conta outros aspectos igualmente importantes antes de considerar qual a arquitectura a utilizar. Um desses aspectos, e que condiciona o uso do dispositivo, é a autonomia – será o uso extensivo das redes móveis para as transferências de dados ou o processamento local que terá um impacto significativo na autonomia do dispositivo? Outros aspectos, também relacionados com o uso das redes móveis, são o custo e a acessibilidade. O uso das redes móveis poderá ser dispendioso para o utilizador, e especialmente no caso desta aplicação, que exige alguma largura de banda, uma vez que os dados transmitidos se tratam de imagens. Esse custo iria variar dependendo o tipo de rede (Wifi, 3G, 4G, etc.) e caso o utilizador esteja no estrangeiro terá custos acrescidos devido ao *roaming*. Por fim, o acesso às redes móveis pode estar condicionado devido a baixa qualidade do sinal ou mesmo falta de cobertura, pelo que a dependência das redes móveis iria, nesse caso, tornar a aplicação inútil.

A solução desenvolvida recorre ao dispositivo móvel para todos os processos, excepto o reconhecimento óptico de caracteres que, devido à duração de processamento no dispositivo, é realizado remotamente.

1.4 Contribuições

As principais contribuições do trabalho desenvolvido nesta dissertação são:

- Concepção e desenvolvimento de uma interface para dispositivos móveis para a digitalização, detecção e classificação automática de recibos em papel que integra realidade;
- Análise e implementação de um sistema para digitalização, detecção e classificação de recibos em papel;
- Concepção, implementação e avaliação de um método de classificação incremental de documentos de texto baseado no classificador Naive Bayes;
- Estudo comparativo de desempenho do sistema, executando todas as tarefas localmente no dispositivo móvel e numa arquitectura cliente-servidor.

O trabalho desenvolvido foi também publicado numa conferência nacional de Informática:

- Miguel Batista, Pedro Morais, Rui Jesus, Pedro Jorge, “FACMOVE: Aplicação para Digitalização, Detecção e Classificação Automática de Recibos em Dispositivos Móveis”, INForum 2012, págs. 172-175, Lisboa, Portugal 2012.

1.5 Organização da Dissertação

Esta secção pretende descrever a organização deste documento. Esta dissertação é composta por um capítulo introdutório, por um capítulo onde é apresentado o trabalho relacionado, por um capítulo sobre o trabalho desenvolvido, por um capítulo de resultados e por um capítulo com as conclusões e trabalho futuro. Os capítulos referidos estão organizados da seguinte forma:

1. Introdução: Apresenta a motivação e os objectivos desta dissertação, enquadra o problema no contexto dos dispositivos móveis, apresenta a solução desenvolvida e as principais contribuições desta dissertação.

2. Trabalho Relacionado: Descreve algumas aplicações relacionadas com esta dissertação dentro das várias áreas que aborda como o reconhecimento óptico de caracteres, o processamento e gestão de recibos ou o processamento de imagem e realidade aumentada. Este capítulo apresenta ainda as tecnologias mais utilizadas para resolver os desafios abordados no desenvolvimento deste trabalho.
3. Digitalização, Detecção e Classificação Automática de Recibos: Explica o processo de desenvolvimento da aplicação, desde a aquisição e detecção de recibos até à classificação dos mesmos.
4. Resultados: Apresenta os resultados obtidos em várias etapas do processamento e classificação de recibos, e discute algumas limitações e as soluções encontradas para ultrapassar as mesmas.
5. Conclusão e Trabalho Futuro: Conclui o documento apresentando as conclusões sobre o desenvolvimento desta dissertação e os resultados obtidos, e apresenta uma perspectiva geral sobre o que poderá vir a ser feito em termos de trabalho futuro para melhorar ou reavaliar certos aspectos desta dissertação.

Trabalho Relacionado

Neste capítulo são apresentados alguns trabalhos relacionados com o sistema desenvolvido nesta dissertação. Em primeiro lugar, são descritos alguns sistemas que se aproximam da nossa proposta; de seguida são apresentadas soluções relacionadas com áreas de conhecimento que estão incluídas no sistema proposto nesta dissertação. As áreas relacionadas são as áreas de reconhecimento óptico de caracteres (OCR), processamento e classificação de recibos, e realidade aumentada. São dados exemplos de algumas aplicações existentes nessas áreas, incluindo a descrição e discussão das suas componentes próximas da nossa proposta.

2.1 Gestão de Recibos em Dispositivos Móveis

Na área de Gestão de Recibos em Dispositivos Móveis existem algumas aplicações disponíveis como, por exemplo, as aplicações Shoeboxed Receipt Tracker and Reader [5] e Expensify [14].

A primeira [5] é uma aplicação em que o utilizador tira uma fotografia do recibo, envia-a para um servidor, sendo o recibo processado até um dia útil após o envio da fotografia. Quando o processamento do recibo estiver completo, este é adicionado à lista de recibos do utilizador, que estará organizada em categorias conforme o tipo do mesmo. Contudo, não é mencionada a forma como se faz o processamento dos recibos. O Shoeboxed Receipt Tracker and Reader permite uma utilização gratuita até 5 recibos por mês.

A aplicação Expensify [14] é semelhante à anterior e é gratuita se o utilizador criar os registos digitais dos recibos manualmente. Caso o utilizador pretenda a funcionalidade de extracção automática de informação e

classificação apenas pode processar 10 recibos por mês, sendo que após este limite o serviço começa a ser cobrado. A funcionalidade de extracção automática de informação extrai o nome da entidade emissora, data do recibo e valor total do recibo, e pode demorar até 10 minutos. Caso a extracção automática de informação falhe, a imagem do recibo é revista por uma pessoa e o resultado é enviado ao utilizador no dia útil seguinte.

A Figura 8 e a Figura 9 mostram os *layouts* principais destas duas aplicações, o Shoeboxed Receipt Tracker and Reader e o Expensify.

As aplicações acima descritas, embora sejam interessantes do ponto de vista de processamento de recibos em papel, divergem da aplicação desenvolvida nesta dissertação uma vez que não dispõem da componente de realidade aumentada e não conseguem fazer o processamento do recibo – extracção de dados e classificação – em tempo útil, ou seja, a duração de todo o processo não é confortável do ponto de vista de utilização por parte do utilizador uma vez que em [5] demora até um dia útil e em [14] pode demorar até 10 minutos, ou caso falhe, até um dia útil.

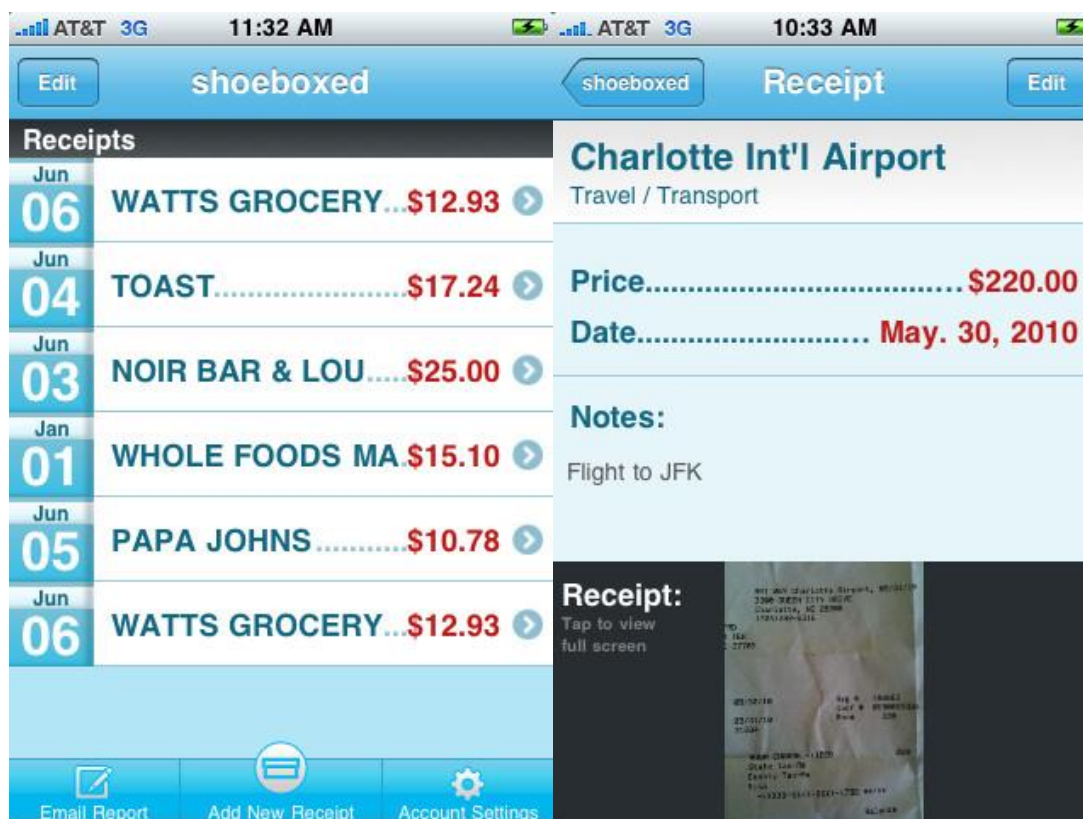


Figura 8 – Shoeboxed Receipt Tracker and Reader.

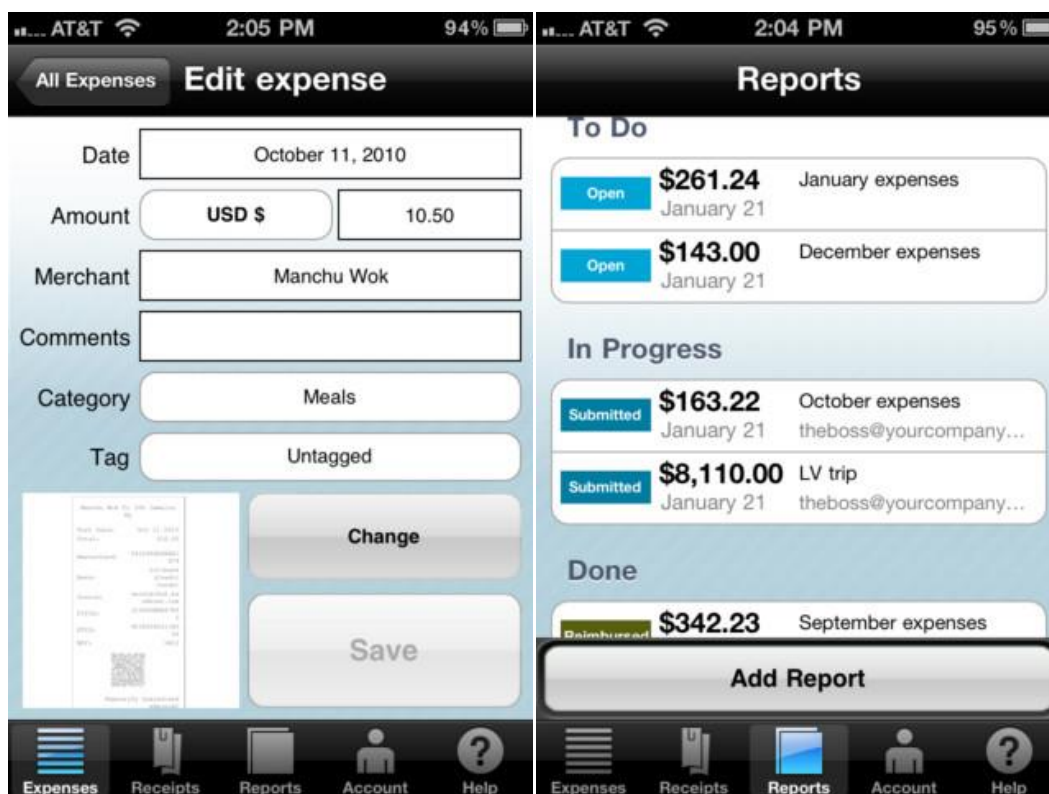


Figura 9 – Aplicação Expensify.

2.2 Reconhecedor Óptico de Caracteres

Uma das componentes mais importantes numa aplicação como a que é proposta neste documento é o Reconhecedor Óptico de Caracteres. Existem várias aplicações de OCR disponíveis para dispositivos móveis que permitem obter imagens com qualidade suficiente, graças às câmaras dos mais recentes *smartphones*. Alguns exemplos dessas aplicações são o Image To Text [15] e o FotoNote [16].

A aplicação Image To Text [15] processa uma imagem captada directamente da câmara do dispositivo ou uma imagem pré-gravada, extraíndo o texto nela contido. A imagem é enviada para um servidor remoto onde será realizado o OCR e, após a conclusão, o resultado é enviado por email para o utilizador (ver Figura 10).

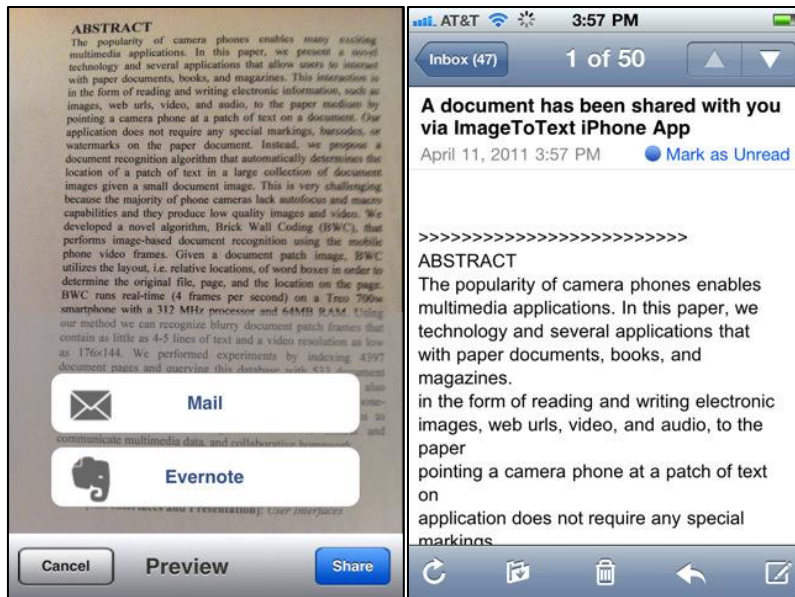


Figura 10 – Image to Text – OCR.

O FotoNote [16] é outra aplicação de OCR que faz processamento de uma imagem para extracção de texto e que permite a exportação do texto detectado. Antes de executar o OCR, esta aplicação transforma a imagem original numa imagem binária, ou seja, uma imagem a preto e branco com 1 bit por pixel. Na Figura 11 apresenta-se um dos *layouts* desta aplicação.

Em ambas as aplicações não é referido qual o software de OCR utilizado, e por isso não é possível saber se é um OCR proprietário ou *open-source*.

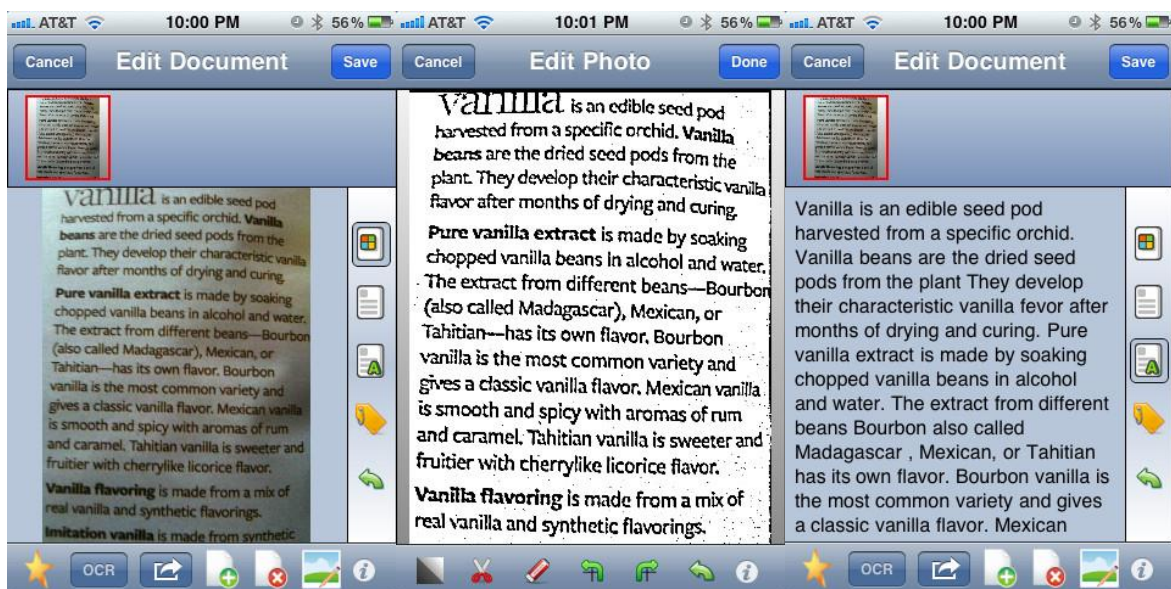


Figura 11 – FotoNote.

2.2 Processamento de Recibos no Servidor versus Cliente

As questões mais críticas aplicações de processamento e classificação de recibos estão relacionadas com o desempenho do OCR e com a detecção do recibo após a digitalização. Por isso, são baseadas na arquitectura cliente-servidor, tal como é sugerido em [17] numa aplicação de reconhecimento de objecto. O cliente apenas serve para tirar a foto. O servidor inclui todo o processamento mais pesado do ponto de vista computacional. Contudo, em algumas circunstâncias pode ser difícil a ligação a um servidor. Em [18] é apresentada uma solução de reconhecimento de objectos que mostra que é possível com os actuais dispositivos móveis criar aplicações que não necessitam de servidor.

No contexto das aplicações de reconhecimento automático de recibos, uma solução que não necessite de servidor terá de incluir uma aplicação de OCR no dispositivo móvel. Existem várias aplicações de OCR disponíveis para dispositivos móveis que permitem obter imagens com qualidade suficiente, graças às câmaras dos mais recentes *smartphones*, como as aplicações descritas anteriormente Image To Text [15] e FotoNote [16].

2.3 Processamento de Imagem de Documentos

Existem algumas aplicações que fazem processamento de imagens de documentos, para o utilizador ficar com uma cópia digital do documento desejado. Dois exemplos deste tipo de aplicação são as aplicações JotNot Scanner [19] e Genius Scan [20]. A Figura 12 e a Figura 13 mostram *layouts* destas duas aplicações. Estas aplicações recorrem a técnicas de processamento de imagem para normalizar a imagem, ou seja, corrigir rotações e efeitos de perspectiva. Na primeira aplicação [19], o utilizador tira uma fotografia ao documento e de seguida selecciona os quatro cantos do mesmo, sendo que a zona seleccionada será processada de forma a obter uma imagem normalizada. A imagem normalizada resultante é colocada à disposição na rede local, num servidor aplicacional contido na aplicação, ao qual o utilizador pode aceder através de um *browser* no computador, e arquivar a imagem do documento.

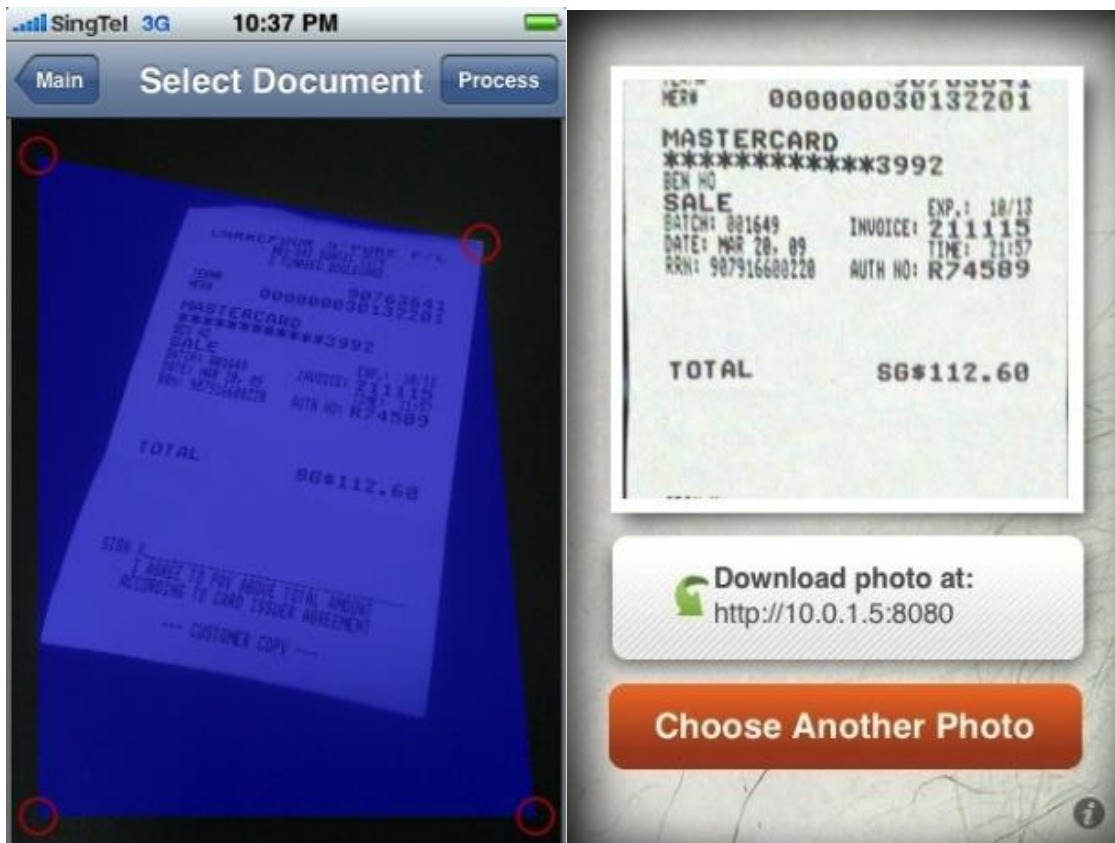


Figura 12 – JotNot Scanner.

A segunda aplicação de processamento de imagem de documentos [20] permite ao utilizador tirar uma fotografia a um documento com a câmara do dispositivo móvel, ou então usar uma imagem guardada no dispositivo. A aplicação faz detecção automática dos contornos do documento e permite ao utilizador corrigir os mesmos (se for necessário) utilizando a mesma técnica utilizada na aplicação JotNot Scanner [19]. Esta aplicação também produz uma imagem normalizada e permite escolher o tipo de tratamento dado à imagem, ou seja, permite obter uma imagem a preto e branco, com melhoramento de contraste ou sem nenhum efeito. Por fim, esta aplicação permite a exportação do documento em formato JPEG e PDF para diferentes destinos como por exemplo, email, Dropbox, Evernote ou Google Drive.

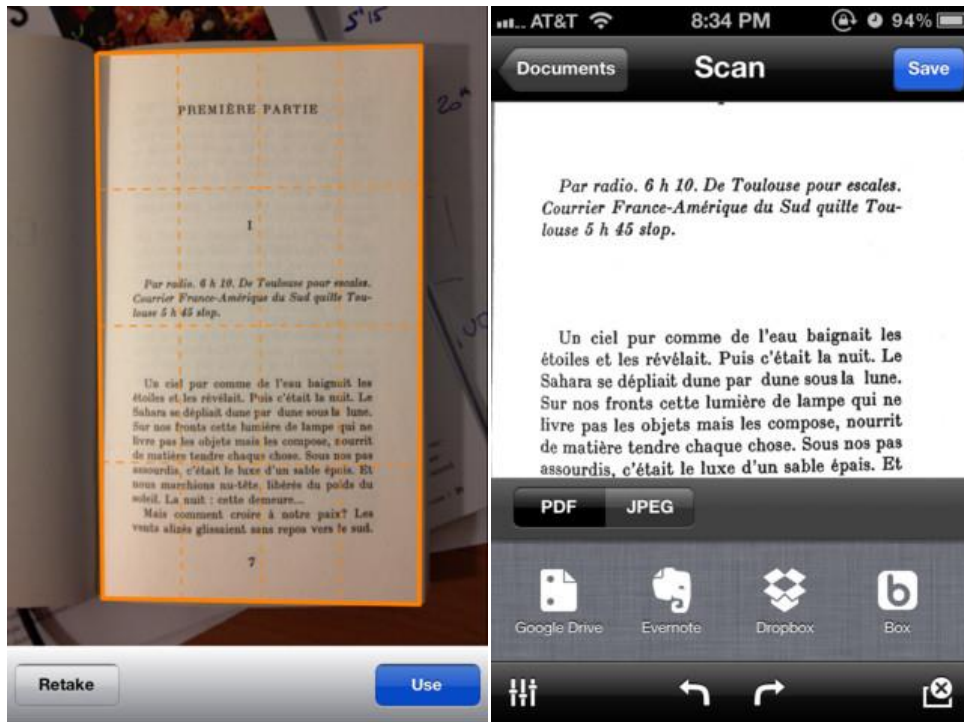


Figura 13 – Genius Scan.

2.4 Realidade Aumentada

A realidade aumentada é um dos campos que têm vindo a crescer, também com a ajuda da evolução dos dispositivos móveis que cada vez mais recorrem a esta técnica. Realidade aumentada [21] é a sobreposição de gráficos gerados por computador sobre imagens do mundo real com o objectivo de dar ao utilizador mais informação acerca daquilo que ele vê naquele instante. Nesta dissertação é utilizada realidade aumentada que mostra texto e gráficos 2D. No entanto, é possível utilizar modelos 3D em dispositivos móveis. Existem aplicações que utilizam objectos 3D complexos em aplicações de realidade aumentada que respondem, por exemplo, à orientação do utilizador.

No mercado existem várias aplicações que usam realidade aumentada. Desde jogos até sistemas de tradução de texto em tempo real e sistemas de visualização de modelos 3D, entre outros. Alguns exemplos de aplicações de realidade aumentada existentes são a aplicação Google Goggles [22], Layar [23] e Word Lens [24].

A primeira [22] detecta e identifica os objectos presentes na imagem, mostrando uma lista com os resultados mais relevantes acerca dos objectos detectados. Inicialmente esta aplicação apenas suportava a pesquisa de objectos em fotografias, mas actualmente suporta a pesquisa de objectos em tempo real, ou seja, pesquisa contínua sem necessidade de tirar uma fotografia. O Google Goggles consegue ainda funcionar como leitor de códigos de barras e códigos QR. Na Figura 14 apresenta-se um exemplo de *layouts* desta aplicação.

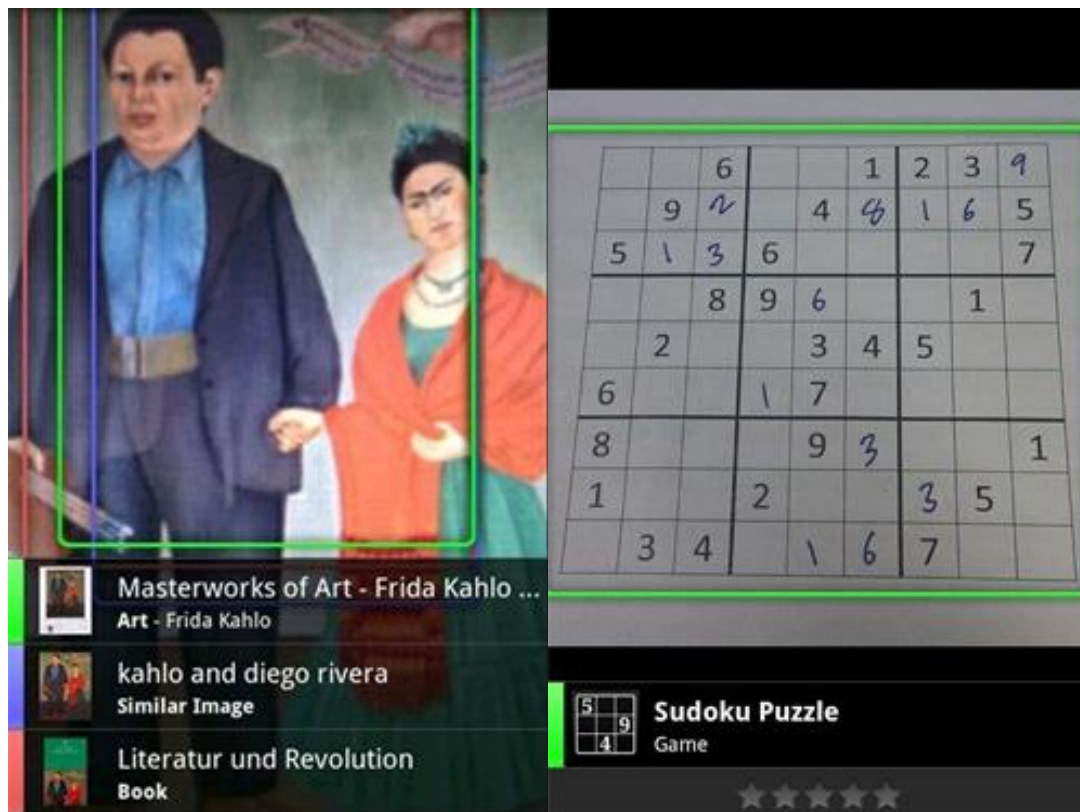


Figura 14 – Google Goggles.

A segunda aplicação, Layar [23], é uma aplicação que utiliza realidade aumentada para dar ao utilizador informação variada acerca de locais de interesse próximos do mesmo. Estas informações podem ser sobre agências bancárias próximas, restaurantes, cafés, caixas multibanco, entre outros. A informação geográfica é uma parte fundamental desta aplicação uma vez que se baseia na localização do utilizador para obter os locais de interesse mais próximos deste. Toda a informação acerca dos locais de interesse próximos é colocada em tempo real sobre a imagem captada pelo dispositivo (ver Figura 15).

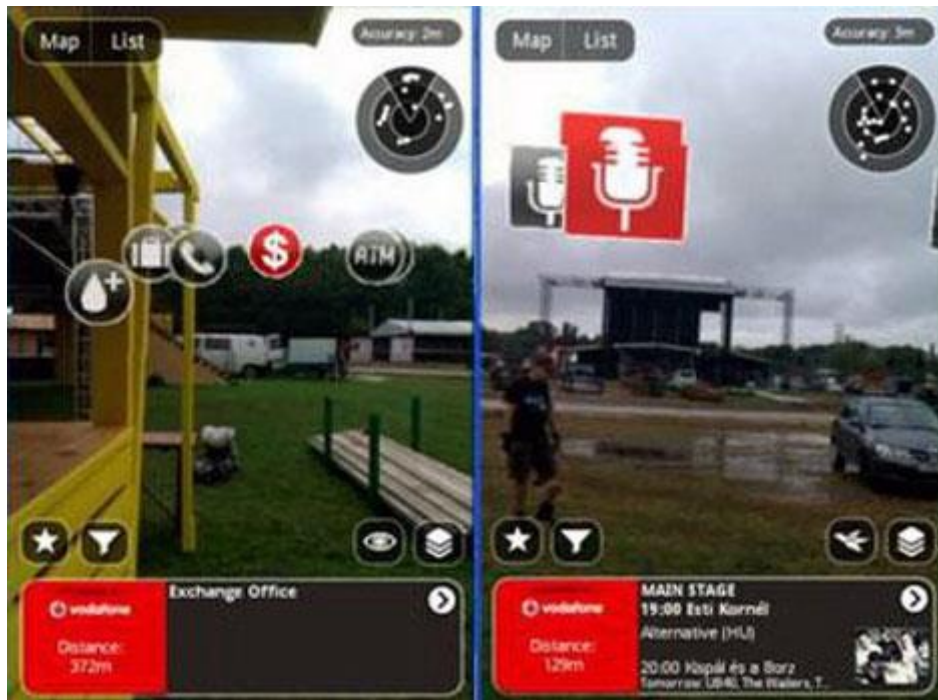


Figura 15 – Layar.

Finalmente, o Word Lens [24] é uma aplicação que permite fazer traduções em tempo real alterando a imagem captada pela câmara com o texto traduzido. A aplicação tem suporte para Inglês, Italiano, Espanhol e Francês. Permite fazer traduções, inverter a ordem das palavras e remover palavras da imagem por completo. A Figura 16 mostra um exemplo da interface desta aplicação.



Figura 16 – Word Lens.

2.5 Tecnologias

Com vista a cumprir os objectivos deste trabalho fez-se um estudo sobre as tecnologias existentes nas diferentes áreas que esta dissertação aborda.

No que diz respeito à área do processamento de imagem, existem várias bibliotecas disponíveis como, por exemplo, o OpenCV [25], os *frameworks* de iOS CoreImage [26] e CoreGraphics [27]. O OpenCV é uma biblioteca de processamento de imagem já bastante conhecida e que agora tem também versões para Android [28] e iOS [29]. Esta biblioteca disponibiliza uma série de conceitos de alto nível, como por exemplo contornos de objectos ou padrões, o que permite um manuseamento e processamento mais fáceis da imagem. Existem versões do OpenCV para ambientes Windows e Linux, implementadas em C e em Python – útil para processamento remoto num servidor. Os *frameworks* CoreImage e CoreGraphics do iOS, embora não tenham tantas funções como o OpenCV, permitem aplicar algumas transformações e filtros a imagens ou processar imagens ao nível do pixel.

Como biblioteca de processamento de imagem a usar para a realização desta dissertação escolheu-se o OpenCV, por permitir um processamento mais completo de imagens do que os *frameworks* de iOS.

Em termos de bibliotecas de reconhecimento óptico de caracteres foi escolhido o OCR Tesseract pelo facto de ser o que apresenta melhor

qualidade. Este OCR é tido em conta como um dos melhores OCR *opensource* existentes. O Tesseract esteve no top 3 dos melhores OCR do Mundo no UNLV AccuracyTest de 1995 [30]. Este OCR suporta imagens binárias, em escala de cinzento e a cores, permite treino para se ajustar a diferentes Línguas e é possível usá-lo em dispositivos móveis e em ambientes Windows e Linux, útil no caso de utilização de um servidor para processamento remoto.

Na área da classificação e aprendizagem automática existem algumas opções como as redes neuronais [31] ou o classificador Naive Bayes [32]. O primeiro [31] apresenta uma capacidade de classificação e uma flexibilidade que se adequa ao problema da classificação de recibos. Existem várias bibliotecas que implementam as redes neuronais em C/C++ como, por exemplo, o NNF2 [33] e o Fast Artificial Neural Network (FANN) [34]. O classificador Naive Bayes [32] segue as propriedades probabilísticas dos dados usando o método da máxima verosimilhança. É necessário treinar o algoritmo com alguns dados de treino. O conjunto de treino poderá ir crescendo à medida que se vão classificando recibos de forma a tornar a classificação melhor e mais abrangente. Uma das bibliotecas existentes para classificação que suportam este método é a biblioteca Bow [35].

Para a realização desta dissertação optou-se por implementar o algoritmo do classificador Naive Bayes pelo facto de ser um algoritmo simples e de as bibliotecas referidas anteriormente para Redes Neuronais e para o classificador Naive Bayes não estarem disponíveis para dispositivos móveis.

Para a realidade aumentada existem algumas bibliotecas, como o AR Toolkit [36], o Mobile AR [37] e o QCAR [38]. O AR Toolkit foi um dos primeiros projectos na área da realidade aumentada e de onde surgiram depois vários projectos. O projecto AR Toolkit parece ter sido abandonado uma vez que não saem novas versões desde 2007. Também o facto do AR Toolkit só estar disponível para Windows, Linux e Mac inviabilizam o seu uso em dispositivos móveis. O Mobile AR foi um dos projectos que nasceu do AR Toolkit e disponibiliza um SDK tanto para iOS como para Android mas tem uma licença paga e assim sendo não se enquadra nos requisitos que se pretendem. Por fim, o QCAR é um SDK desenvolvido pela Qualcomm e tem suporte directo para iOS sob a forma de uma biblioteca em Objective-C.

Para a realização deste trabalho optou-se pela utilização do *framework* de iOS CoreGraphics para criar uma realidade aumentada 2D de baixa complexidade mas, para trabalho futuro, a biblioteca QCAR será a melhor opção para desenvolver uma realidade aumentada mais complexa.

Digitalização, Detecção e Classificação Automática de Recibos

Neste capítulo, descrevem-se os diferentes blocos que constituem a aplicação para a digitalização, detecção e classificação automática de recibos em dispositivos móveis, desde o processamento inicial da imagem até à classificação da informação obtida do recibo.

3.1 Diagrama de Blocos

Como descrito anteriormente, os dispositivos móveis incluem uma câmara digital cada vez com maior resolução, permitindo a sua utilização para realizar a digitalização do recibo. Para melhorar o desempenho do sistema, quer em termos de detecção do texto incluído no recibo, quer em termos de tempo de processamento, é desejável que a imagem adquirida contenha somente a área de interesse, definida numa zona rectangular. Assim, são utilizadas técnicas de processamento de imagem para detectar e normalizar o recibo digital (Aquisição e Detecção).

Após este passo, é aplicado um algoritmo de reconhecimento óptico de caracteres (OCR – *Optical Character Recognition*) para a extracção do texto incluído no recibo. Este procedimento é crítico uma vez que todos os processos seguintes dependem do desempenho desta detecção. O texto obtido irá ser utilizado para a extracção de informação relacionada com o recibo, como por

exemplo, o Número de Identificação Fiscal (NIF) da entidade que o emitiu, o seu valor total ou o valor do Imposto sobre o Valor Acrescentado (IVA).

Esta aplicação também classifica o recibo consoante a finalidade do produto adquirido, como por exemplo, restauração, combustível ou saúde.

Este procedimento também é realizado com base nas palavras extraídas pelo OCR. A informação extraída é adicionada à imagem digital do recibo permitindo criar uma entidade multimédia com metadados (Extracção de Informação). Este registo pode ser visualizado recorrendo a realidade aumentada, sobrepondo elementos virtuais que representam os metadados à imagem digital do recibo, adquirida no instante actual ou armazenada (Interface Gráfica). Paralelamente, os registos que vão sendo adquiridos podem ser organizados, permitindo por exemplo, a sua consulta, o cálculo de somas parciais por categoria ou a soma total (Gestão). As secções seguintes descrevem cada bloco apresentado anteriormente e ilustrados na Figura 17.

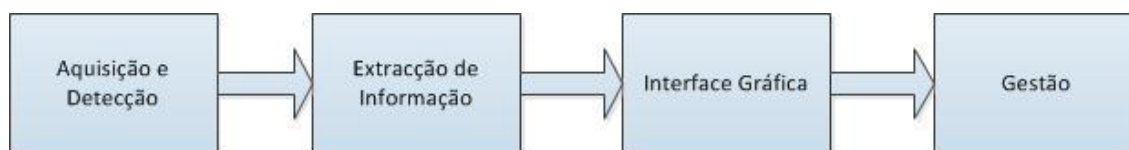


Figura 17 – Diagrama de Blocos do Sistema Simplificado.

3.2 Aquisição e Detecção de Recibos

A finalidade deste bloco consiste na transformação da informação visual constante no recibo em papel numa imagem digital contendo a zona útil do recibo onde se encontra o texto, como exemplifica a Figura 18. Esta operação é necessária para determinar a localização e a perspectiva com que se adquiriu o recibo, útil para a interface gráfica e para a detecção do texto.

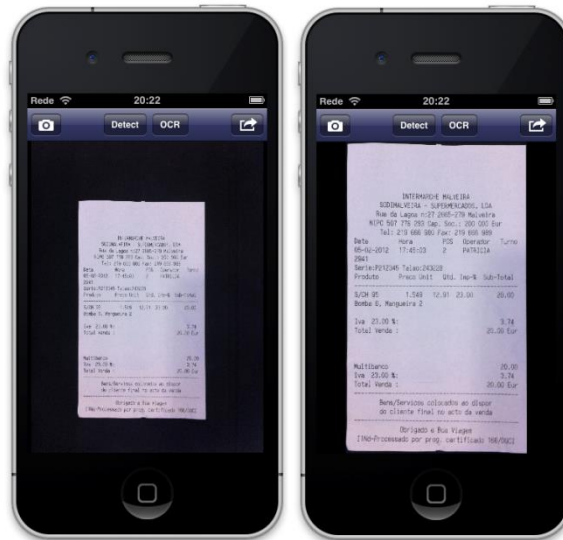


Figura 18 – Exemplo de Detecção de Recibo.

Em primeiro lugar é realizado um escalamento à imagem original para uma dimensão inferior e uma conversão para níveis de cinzento. Este passo permite limitar o tempo de processamento em virtude de se processarem imagens com menor resolução e somente com uma componente de intensidade. Dadas as características geométricas do recibo, normalmente de forma rectangular, devido ao efeito de perspectiva, é projectado no plano de imagem como um quadrilátero. Assim, optou-se por aplicar o algoritmo da Transformada de Hough [39] para a detecção das rectas que delimitam a área de interesse (ROI – *Region of Interest*). Com base na intersecção destas rectas é possível definir o contorno exterior do recibo. Aplicando uma transformação homográfica [40] é possível compensar o efeito de perspectiva e obter uma imagem do recibo rectangular que contém a informação útil necessária.

O diagrama de blocos da Figura 19 resume o conjunto de operações realizadas no processo de Aquisição e Detecção.

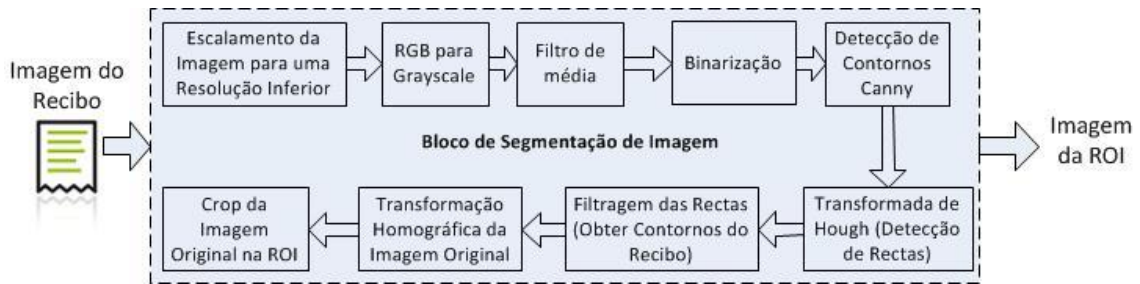


Figura 19 – Diagrama de blocos da Aquisição e Detecção do Recibo.

Após o escalamento da imagem original para uma dimensão inferior e de transformar esta numa imagem de níveis de cinzento, é aplicado um filtro de média para eliminar ruído que possa existir na imagem (ver Figura 20 à esquerda). De seguida é feita a binarização da imagem com limiar global para transformar a mesma numa imagem de 1 bit/pixel (ver Figura 20 à direita) que será utilizada pelo algoritmo de detecção de contornos de Canny [41].

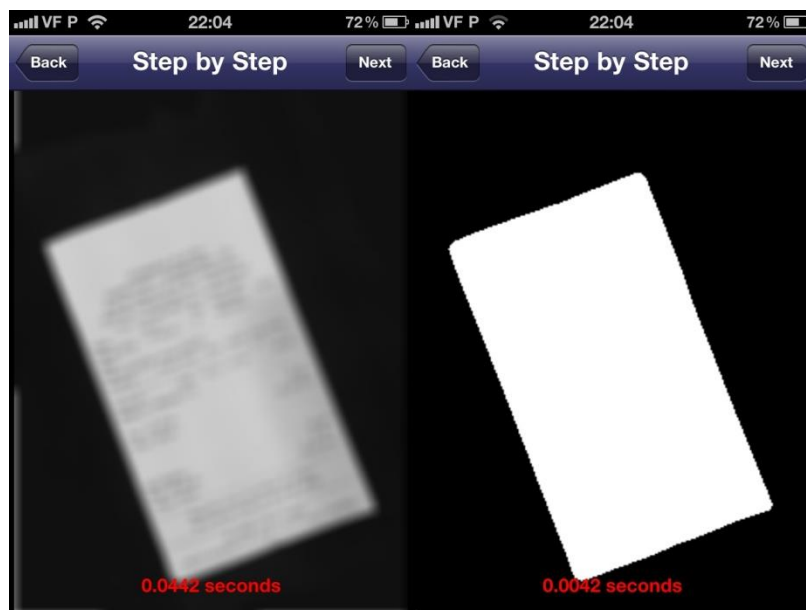


Figura 20 – Filtragem com filtro de média (à esquerda); Binarização (à direita).

Utilizando o algoritmo de detecção de contornos de Canny obtém-se uma imagem apenas com os contornos existentes na imagem (ver Figura 21 à esquerda). Esta imagem pode então ser processada utilizando a Transformada de Hough [39], para detecção de rectas. Adicionalmente, é realizada uma filtragem às rectas encontradas pelo algoritmo, de forma a separá-las em

verticais e horizontais, e foram removidas as rectas muito próximas umas das outras e as que não são consideradas paralelas. Das rectas resultantes desta filtragem foram escolhidas as que ficam nos extremos. Esta filtragem considera que os recibos são rectangulares (ver Figura 21 à direita).

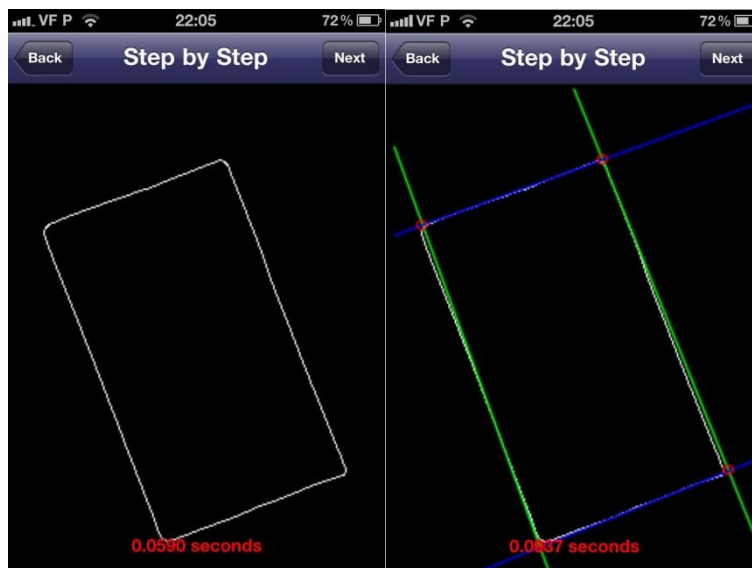


Figura 21 – Detecção de contornos de Canny (à esquerda); Detecção dos limites do recibo com base na Transformada de Hough (à direita).

Obtendo as rectas que definem os contornos do recibo, são calculados os pontos de intercepção dos mesmos para obter os vértices da área de interesse. Como todo o processamento até aqui foi efectuado sobre uma imagem com dimensão inferior é necessário transformar os pontos obtidos para a escala da imagem original. Estes pontos são usados para efectuar a transformação homográfica [40] sobre a imagem original, que irá compensar o efeito de perspectiva ou rotação do recibo. Finalmente, é feito um corte da imagem resultante da transformação homográfica, de forma a obter uma imagem apenas com o recibo, a qual será utilizada nos processos seguintes e será guardada juntamente com os dados extraídos do recibo. A Figura 22 mostra um exemplo de uma transformação homográfica e o corte posterior.

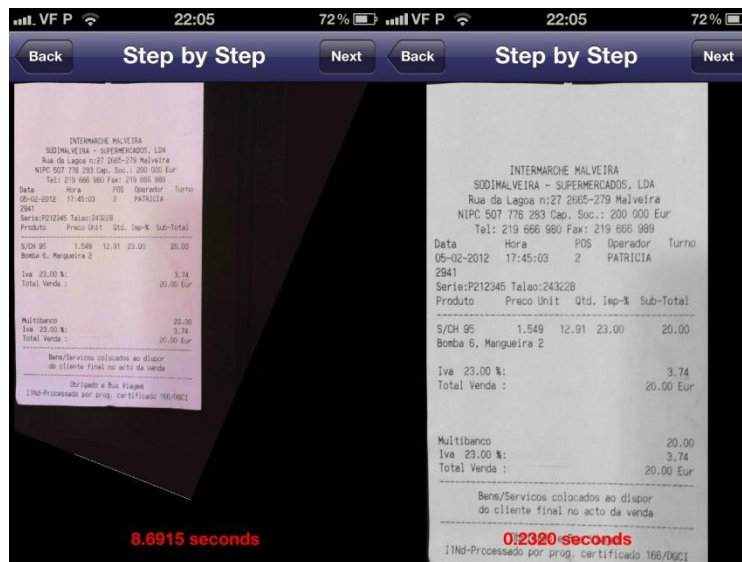


Figura 22 – Transformação homográfica (à esquerda); *Crop* do recibo (à direita).

3.3 Extracção de Informação

A partir da imagem do recibo, pretende-se extrair a informação útil para o objectivo do sistema implementado. Optou-se neste trabalho por considerar a informação do NIF da entidade que emitiu o recibo e o valor total do mesmo. Contudo, de forma análoga, pode-se extrair outros campos do recibo, com por exemplo, o IVA ou a sede da entidade. O conjunto de palavras extraído será utilizado para a classificação do recibo. A Figura 23 resume os blocos constituintes deste passo.

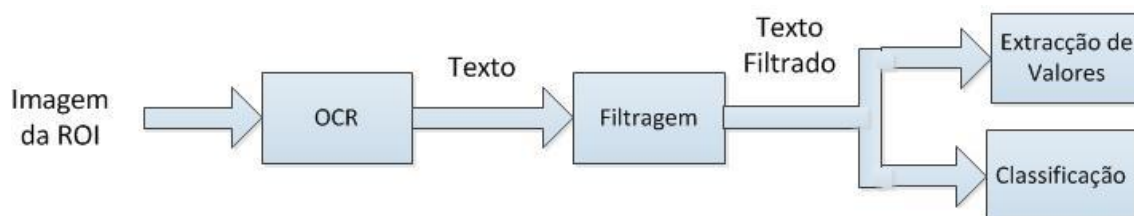


Figura 23 – Diagrama de blocos da Extracção de Informação.

Em primeiro lugar é aplicado um reconhecimento óptico de caracteres, e em seguida é realizada uma filtragem das palavras detectadas, formando um conjunto de palavras característico do recibo. Este conjunto é utilizado para a

extração dos valores pretendidos do recibo e para realizar a sua classificação. À medida que são extraídos os conjuntos de palavras dos vários recibos, o sistema vai actualizando um dicionário de palavras necessário para a classificação. As subsecções seguintes descrevem cada um destes blocos com mais detalhe.

3.3.1 Reconhecimento Óptico de Caracteres



Figura 24 – Diagrama de blocos do Reconhecimento Óptico de Caracteres.

Para a realização do reconhecimento automático de caracteres é utilizada a aplicação de código aberto *Tesseract* [30], inicialmente desenvolvido pela *HP Labs* e actualmente mantido pelo *Google*. Esta aplicação é considerada um dos OCR de código aberto com melhor desempenho e tem como interface de entrada uma imagem digital, gerando como saída um ficheiro de texto contendo os caracteres e outros símbolos reconhecidos. A Figura 24 mostra um exemplo desta interface. O conjunto de caracteres extraídos pelo OCR é filtrado, de forma a limitar o número de palavras a incluir no dicionário comum ao sistema. Este processo inclui a eliminação de palavras com menos de 3 caracteres, assim como símbolos e pontuações.

3.3.2 Extração de Valores

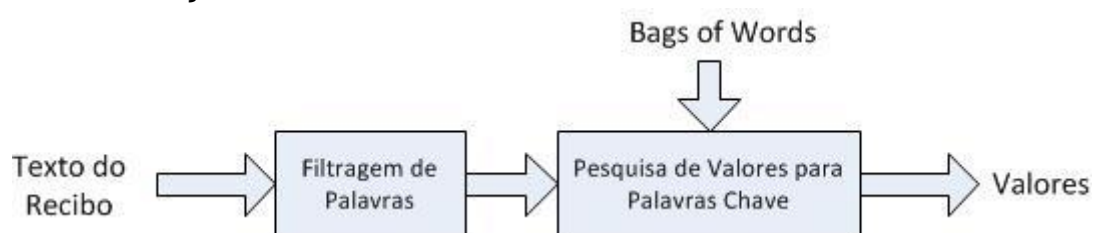


Figura 25 – Diagrama de blocos da Extração de Valores.

Com base num conjunto de recibos diferentes foi possível concluir que existe um conjunto de palavras associadas a conceitos como o valor total do recibo ou o número de contribuinte da entidade emissora do recibo. Assim, de forma a efectuar a extracção desses valores do recibo, é construída uma colecção de palavras (*bag of words*) que representa cada campo a detectar. Cada colecção é formada pelas várias palavras com que cada item poderá ocorrer no recibo. Por exemplo, para representar o Número de Identificação Fiscal podem ocorrer as palavras “NIF”, “NIPC” (Número de Identificação de Pessoa Colectiva) ou “n cont” (Número de Contribuinte).

O processamento de cada recibo envolve a extracção do seu conjunto de palavras e, para cada um dos itens que se pretende detectar, é verificada a ocorrência das palavras existentes no seu *bag of words* (ver Figura 25), sendo esta verificação auxiliada pela utilização da distância de Hamming [42] uma vez que, devido a erros de detecção do OCR, os termos pelos quais se procura poderão não ser exactamente iguais ao esperado. Neste caso é utilizada uma distância de Hamming de 2 (ver Figura 26). Em seguida, é extraído o valor desse campo, tendo em atenção a sua sintaxe, como por exemplo, o NIF é composto por 9 dígitos numéricos. Esta informação é adicionada ao registo informático do recibo.

Termo a procurar:	total	total
Termo OCR:	t0ta1	tanto
Distância de Hamming:	2	4
	✓	✗

Figura 26 – Exemplo de distância de Hamming entre termos.

A Figura 27 contém dois exemplos de recibos, onde estão assinalados os dados a serem extraídos. É possível verificar que a palavra-chave (por exemplo, “NIPC” ou “Total”) é seguida do valor respectivo.

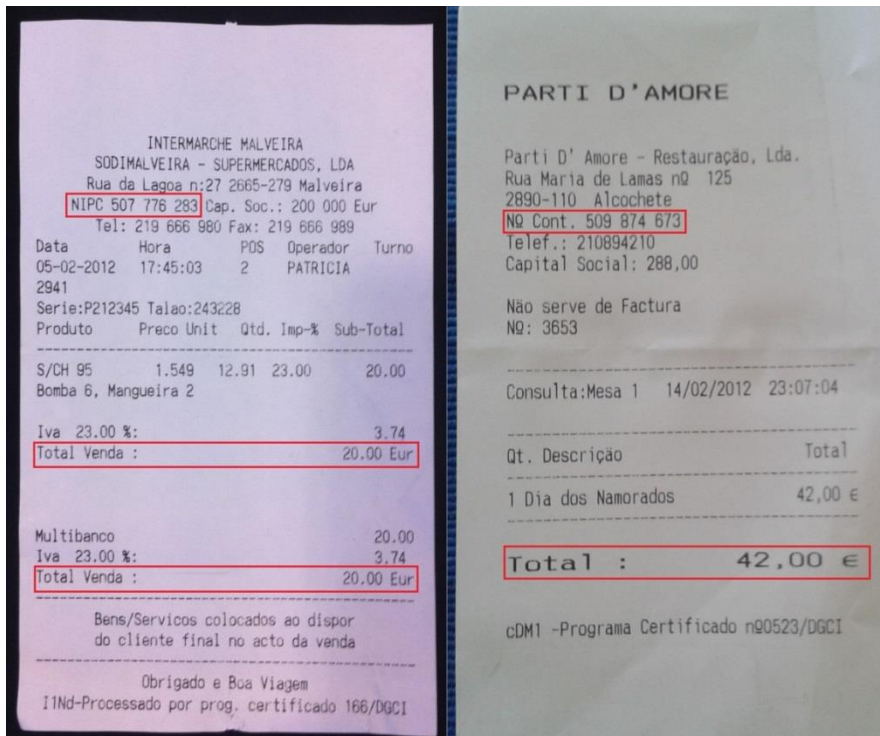


Figura 27 – Exemplos de recibos e dados a extrair.

3.4 Classificação

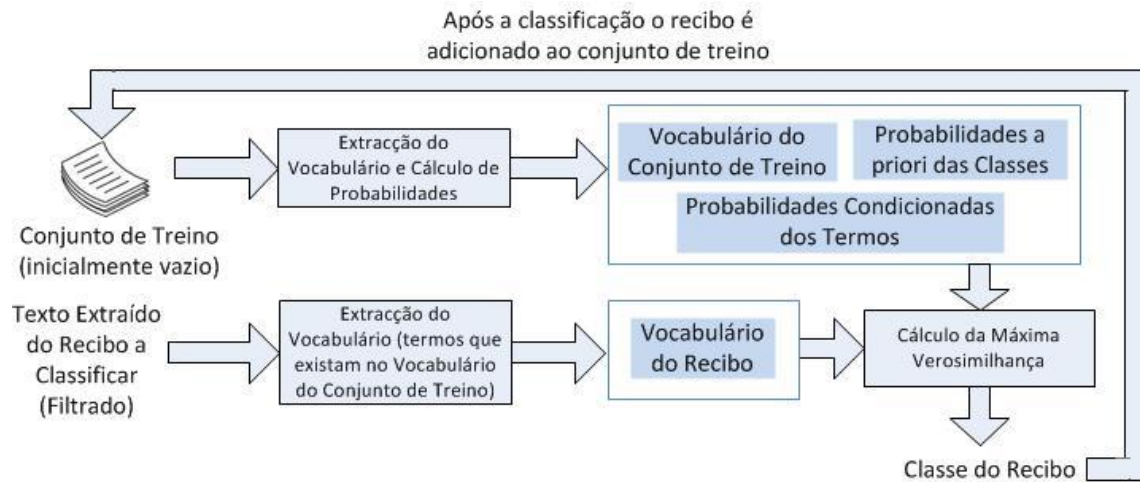


Figura 28 – Diagrama de blocos da Classificação.

Uma aplicação de gestão de recibos deve permitir a organização destes em grupos ou classes com algum significado semântico, como, por exemplo, recibos relacionados com alimentação ou combustível. Assim, é implementado um algoritmo de classificação automática dos recibos que são adquiridos pela aplicação. A Figura 28 resume este processo.

A classificação é realizada com base no algoritmo Naive Bayes [32]. É considerado que cada recibo é composto pelo conjunto de palavras extraídas pelo OCR e que ocorrem de forma independente. É utilizada uma aprendizagem incremental, uma vez que o conjunto de treino (inicialmente vazio) vai aumentando à medida que os recibos vão sendo processados.

O algoritmo de treino é o seguinte:

- a) É extraído o *Vocabulário* (conjunto de palavras, pontuações, símbolos distintos) presente no texto total do conjunto de treino.
- b) Calcula-se a probabilidade *a priori* das classes, segundo a equação (1):

$$P(\text{Classe}_j) = \frac{N^{\circ} \text{ Documentos da Classe}_j}{N^{\circ} \text{ Total de Documentos}} \quad (1)$$

- c) Calcula-se a probabilidade condicionada de todos os termos do vocabulário:
 - a. Concatena-se o texto de todos os documentos do conjunto de treino que pertencem à Classe_j, obtendo o *Texto_j*.
 - b. Calcula-se *n*, o número total de elementos distintos em *Texto_j*.
 - c. Para cada palavra W_k presente no *Vocabulário*:
 - i. Calcula-se n_k , o número de vezes que W_k ocorre em *Texto_j*.
 - ii. Calcula-se a probabilidade da palavra pertencer à classe, utilizando a equação (2):

$$P(W_k | \text{Classe}_j) = \frac{n_k + 1}{n + |\text{Vocabulário}|} \quad (2)$$

- d. Repete-se para as restantes classes

Após o cálculo das probabilidades, o algoritmo está pronto para classificar um documento novo. Para isso é apenas necessário executar dois passos:

- a) Obtém-se a lista de palavras do novo documento que fazem parte do *Vocabulário*, representado por *Termos*.
- b) Calcula-se a máxima verosimilhança, V_{NB} , para obter a classe mais provável a que o novo documento pertence, segundo a equação (3):

$$V_{NB} = \mathit{arg} \max_{Classe_j \in Classes} P(Classe_j) \prod_{W_k \in Termos} P(W_k | Classe_j) \quad (3)$$

Quando é adicionado um novo recibo ao conjunto de treino, é necessário treinar novamente o classificador uma vez que o vocabulário cresce e as probabilidades mudam.

Resultados

Durante a realização deste trabalho foram realizados testes cujos resultados ajudaram a tomar decisões e a convergir para a solução final. Neste capítulo são apresentados os resultados dos testes realizados, e que foram necessários para decidir qual a arquitectura de sistema a utilizar, assim como os testes finais realizados para avaliar o desempenho do sistema desenvolvido. Os testes foram feitos às principais componentes deste trabalho: processamento de imagem e realidade aumentada, OCR e classificação.

4.1 Características dos Dados

Os testes foram realizados utilizando um conjunto de 7 recibos de diferentes dimensões e com diferentes orientações em relação à câmara. Estes recibos foram colocados sobre uma superfície uniforme, sem padrões e com uma cor contrastante em relação ao recibo. Os recibos utilizados pertencem a uma variedade de categorias, como por exemplo, recibos de supermercado, combustível, saúde e vestuário.

4.2 Ambiente de Testes

O dispositivo utilizado para a realização deste trabalho e dos seguintes testes foi um iPhone 4, com sistema operativo iOS 4.2.1, câmara de 5 megapixéis (2592 x 1936 pixéis), 512MB de RAM e processador ARM Cortex-A8 *single core* a 800MHz. A rede utilizada foi uma rede WI-FI baseada em ADSL, com uma velocidade média de *upload* de 90,19 KB/s e uma velocidade média de *download* de 1443,63 KB/s.

4.3 Tempos de Execução

Para avaliar a usabilidade do trabalho realizado começamos por apresentar o tempo total necessário para processar recibos. A medição desse tempo foi efectuada desde o processo de aquisição do recibo (ver Figura 29) até à sua classificação (ver Figura 30 e Figura 31).



Figura 29 – Aquisição do recibo com realidade aumentada.



Figura 30 – Resultado da extracção de valores e classificação do recibo com possibilidade de edição pelo utilizador.

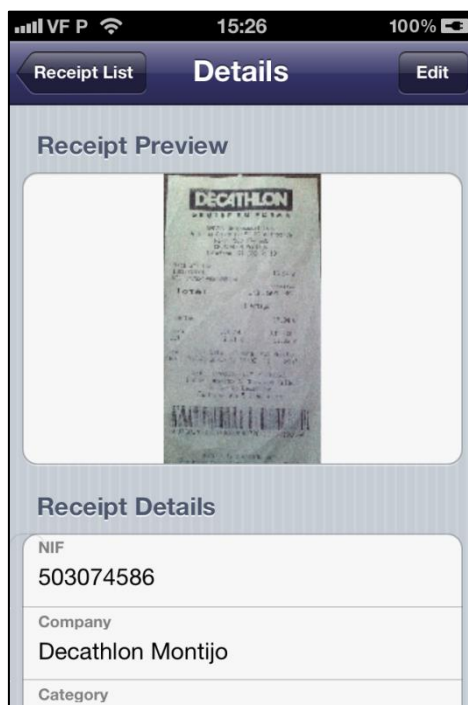


Figura 31 – Visualização do resultado antes de guardar o recibo no dispositivo móvel.

Após a implementação do trabalho totalmente baseado em processamento local, verificou-se que o processamento de recibos não era efectuado em tempo útil e como tal a aplicação era prejudicada em termos de

usabilidade para o utilizador. A Tabela 1 contém a duração média de cada processo envolvido no processamento do conjunto de 7 recibos.

Processo	Duração média (ms)
Escalamento para dimensão inferior	30
Conversão para níveis de cinzento	13
Filtragem com filtro de média	54
Binarização	14
Detecção de contornos de Canny	67
Transformada de Hough e filtragem de rectas	154
Cálculo da matriz de transformação homográfica	21
Transformação homográfica na imagem original	10000
Corte da imagem original na zona do recibo	82
OCR	15000 a >120000
Filtragem do texto e extracção de valores	502
Classificação	61
Total	25998 a >130998

Tabela 1 – Duração média de execução de cada processo.

Com base nos resultados obtidos e apresentados na Tabela 1, verificou-se que os processos responsáveis pela elevada duração total de processamento dos recibos são o OCR e a Transformação Homográfica da imagem original. Assim, concluiu-se que seria necessário encontrar alternativas para reduzir a duração dos processos mais morosos, uma vez que se considerou que um tempo de processamento de recibos entre aproximadamente 26 segundos e mais de 2 minutos não era aceitável para a aplicação desejada.

4.4 OCR

O factor mais limitativo para a usabilidade da aplicação desenvolvida foi o tempo que o OCR demorou a efectuar o reconhecimento de caracteres no dispositivo. Para cada um dos recibos do conjunto de teste foram realizadas avaliações para medir o tempo de processamento do OCR. Para cada um dos

recibos foram efectuados testes com 3 tipos de imagens: cores, níveis de cinzento e binária. Em cada um dos tipos de imagem foram realizados testes com 6 dimensões horizontais diferentes dos recibos, de 1000 a 500 píxeis de largura. Os resultados são apresentados na Tabela 2.

Operação	Duração Média (segundos)
OCR em Imagens a Cores	36.25
OCR em Imagens em níveis de cinzento	32.05
OCR em Imagens Binárias	37.73

Tabela 2 – Duração média do OCR no dispositivo móvel.

A duração média superior a 30 segundos para a execução do OCR no dispositivo móvel, considerada prejudicial para a usabilidade da aplicação, levou à necessidade de procurar alternativas.

A alternativa encontrada foi delegar este processo para um servidor remoto. Após uma breve pesquisa, foram encontradas soluções *Software as a Service (SaaS)* para OCR, ou seja, *Text Recognition as a Service*. Alguns exemplos destes serviços são o *OCR-IT* [43] e o *ABBY Cloud OCR* [44], mas são serviços pagos e embora tenham uma vertente gratuita, este tem uma utilização limitada e é baseada em filas de espera. Foi feito um pequeno teste com o *ABBY Cloud OCR* no dispositivo móvel utilizando a vertente gratuita, mas esta opção foi abandonada quando o tempo em fila de espera ultrapassou os 2 minutos. Embora possa ser uma solução interessante na sua vertente paga, não é a solução pretendida para este trabalho.

Considerando os testes anteriores, optou-se pela criação de um servidor dedicado para o efeito. Assim, foi criada uma instância gratuita no Amazon EC2 [45] com Sistema Operativo Linux. Nesta instância foi instalado o OCR Tesseract para desktop e foi criado um servidor Web que recebe uma imagem por HTTP, executa o OCR sobre essa imagem e responde com o texto extraído. A Tabela 3 apresenta os resultados do teste efectuado utilizando o servidor e com as mesmas imagens do teste do OCR no dispositivo móvel.

Operação	Duração Médio (segundos)
OCR em Imagens a Cores	1.26
OCR em Imagens em níveis de	1.20

cinzento	
OCR em Imagens Binárias	1.41

Tabela 3 – Duração média do OCR no servidor.

A análise dos resultados do teste do OCR no servidor realça que esta solução apresenta grandes melhorias em relação a efectuar o OCR localmente. O tempo de processamento do OCR diminuiu aproximadamente 30 vezes em relação ao tempo de processamento local. Contudo, a dimensão do ficheiro a enviar para o servidor e a largura de banda da ligação também devem ser consideradas para a determinação do tempo total efectivo da realização do OCR remotamente. A Tabela 4 apresenta a dimensão média e duração média de transmissão das imagens para o servidor. Os testes efectuados a estas variáveis mostram que mesmo com uma ligação com baixa largura de banda conseguem-se melhores tempos do que se o processamento for feito localmente.

	Dimensão média do ficheiro (KB)	Duração média de transferência (segundos)
Imagem a Cores	373.71	4.14
Imagem em níveis de cinzento	358.54	3.98
Imagem Binária	236.36	2.62

Tabela 4 – Dimensão média do ficheiro e duração média de transferência.

Em relação à qualidade do texto detectado pelo OCR nos recibos de teste, foi realizada uma análise relativa a cada uma das dimensões horizontais testadas. Os resultados obtidos são apresentados nas tabelas Tabela 5, Tabela 6 e Tabela 7, sendo que no eixo vertical estão as dimensões horizontais dos recibos e no eixo horizontal está a qualidade do texto detectado pelo OCR, variando entre boa, satisfatória e má. Esta classificação foi dada após avaliar o texto detectado em cada recibo, sendo que boa qualidade corresponde a texto em que existem poucos erros e é possível detectar pelo menos um dos valores extraídos (total ou número de contribuinte), qualidade satisfatória corresponde a texto com alguns erros e em que apenas é possível detectar um dos valores

extraídos, e finalmente má qualidade corresponde a texto com muitos erros em que não é possível detectar nenhum dos valores extraídos. Como referido anteriormente, foram utilizados os 7 recibos de teste para verificar a qualidade do texto obtido pelo OCR em cada uma das dimensões horizontais do recibo.

	Boa	Satisfatória	Má
1000 píxeis	3	3	1
900 píxeis	4	2	1
800 píxeis	2	4	1
700 píxeis	2	4	1
600 píxeis	1	4	2
500 píxeis	0	3	4

Tabela 5 – Qualidade do texto detectado pelo OCR em imagens a cores por dimensão horizontal do recibo.

	Boa	Satisfatória	Má
1000 píxeis	2	4	1
900 píxeis	2	4	1
800 píxeis	1	5	1
700 píxeis	3	3	1
600 píxeis	2	3	2
500 píxeis	0	4	3

Tabela 6 – Qualidade do texto detectado pelo OCR em imagens em níveis de cinzento por dimensão horizontal do recibo.

	Boa	Satisfatória	Má
1000 píxeis	1	3	3
900 píxeis	2	2	3
800 píxeis	0	3	4
700 píxeis	0	4	3
600 píxeis	0	2	5
500 píxeis	0	1	6

Tabela 7 – Qualidade do texto detectado pelo OCR em imagens binárias por dimensão horizontal do recibo.

Como é possível verificar pelas tabelas Tabela 5 a Tabela 7, as imagens a cores e em níveis de cinzento, com resoluções horizontais do recibo de 1000 píxeis e 900 píxeis, são as que apresentam melhores resultados. Assim, optou-se por utilizar a imagem a cores com 1000 píxeis de resolução horizontal, pois além de ter obtido uma boa qualidade de texto, permite processar mais

facilmente recibos com margens maiores em que o texto está mais centrado, como exemplificado na Figura 32, para os quais é necessária uma resolução mais alta para que o OCR não tenha dificuldades na detecção de texto. Segundo a documentação do OCR Tesseract, a dimensão das letras, mais propriamente a dimensão da letra 'x' minúscula, deverá ter no mínimo 20 píxeis de altura.

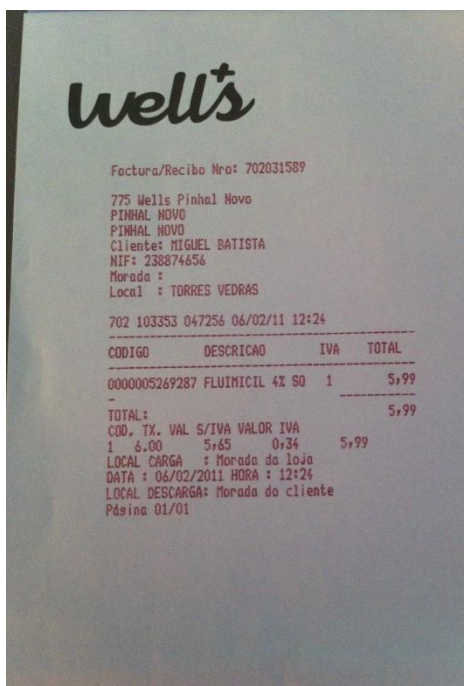


Figura 32 – Exemplo de recibo com margens grandes.

Apesar de ser desejável enviar o mínimo de dados possível pela rede, tanto por motivos de tempo como de custo, verificou-se que devido à utilização de compressão JPEG a diferença de dimensões de ficheiros entre a imagem a cores e a imagem em níveis de cinzento não são muito significativas uma vez que, segundo a Tabela 4, esta é, de apenas 15,17 KB, ou seja, uma diferença de aproximadamente 4%, em média.

Assim, as razões pelas quais se optou por utilizar a imagem a cores são o facto da dimensão média da imagem a cores não ser significativamente superior à da imagem em níveis de cinzento, e da qualidade do texto na imagem a cores ser em média melhor que na imagem em níveis de cinzento (ver Tabela 5 e Tabela 6). A qualidade do texto detectado é de grande

importância para os blocos seguintes, nomeadamente o de extracção de valores e o de classificação.

4.5 Transformação Homográfica

Durante o processamento da imagem do recibo é necessário realizar uma transformação homográfica [40] para compensar possíveis efeitos de perspectiva e também para rodar o recibo caso este não esteja alinhado com a câmara. A biblioteca de processamento de imagem utilizada foi o OpenCV [25]. Esta biblioteca inclui esta funcionalidade, mas o processamento desta transformação no dispositivo foi, em média, de 10 segundos. Para tentar minimizar este tempo tentaram-se outras abordagens.

A primeira abordagem foi utilizar o *framework* de computação gráfica do próprio iOS, o Core Graphics. Este *framework* permite a aplicação de transformações afins, e embora uma transformação homográfica não seja uma transformação afim, realizou-se um pequeno teste para perceber se haveria uma grande diferença em relação aos resultados do OpenCV. A duração média da transformação afim foi cerca de 5 segundos, conseguindo assim reduzir para metade o tempo de processamento, mas a qualidade da transformação era pior que a do OpenCV, uma vez que o recibo não ficava distorcido da mesma forma, havendo situações em que partes do recibo ficavam fora da imagem. A qualidade do texto detectado pelo OCR numa imagem transformada com esta aproximação também era muito pior do que numa imagem transformada pelo OpenCV, havendo casos em que o texto era irreconhecível. Esta abordagem apenas permite realizar transformações afins porque só é possível definir os parâmetros de rotação, escalamento e translação da matriz de transformação como mostra a Figura 33, ao contrário da matriz de transformação do OpenCV em que os parâmetros assinalados na Figura 33 são diferentes de zero.

$$\begin{bmatrix} a & b & 0 \\ c & d & 0 \\ t_x & t_y & 1 \end{bmatrix}$$

Figura 33 – Matriz de transformação afim do Core Graphics.

A segunda abordagem foi utilizar uma biblioteca de processamento de imagem *open source* para iOS [46], baseada em OpenGL que utiliza o poder computacional do GPU para efectuar o processamento. Nesta biblioteca é possível realizar transformações não afins, recorrendo a matrizes de transformações 3D que permitem definir todos os parâmetros necessários, ao contrário da abordagem anterior. A média de tempo de processamento da transformação usando esta abordagem foi de 3 segundos, mas a imagem resultante assim como o texto extraído pelo OCR continuam a ser piores do que usando o OpenCV.

Após a exploração das alternativas acima descritas optou-se por manter a utilização do OpenCV para esta operação, uma vez que apenas com esta ferramenta se conseguiu realizar transformações homográficas que deram origem a imagem normalizadas dos recibos, correctamente centrados e com os efeitos de perspectiva e rotação eliminados.

4.6 Classificador

Para avaliar o desempenho do classificador Naive Bayes efectuaram-se alguns testes iterativos utilizando 6 recibos de 4 classes distintas: Combustível, Restauração, Saúde e Supermercado. Os testes consistiram no seguinte exercício:

- Inicialmente foi classificado correctamente (através de intervenção do utilizador) um recibo de cada classe.
- De seguida, foi testado um recibo de cada classe e verificaram-se os resultados.
- Após o teste anterior, fez-se outra iteração classificando correctamente 2 recibos de cada classe.
- Novamente, testou-se um recibo de cada classe e verificaram-se os resultados.
- Repetiu-se o mesmo exercício adicionando mais um recibo de cada classe a cada nova iteração e classificando um recibo de cada

classe para testar o classificador, até a probabilidade de erro convergir.

Nas tabelas Tabela 8 a Tabela 12 apresentam-se as matrizes de confusão relativas a cada uma das iterações realizadas, cujas linhas correspondem à classe verdadeira e as colunas correspondem à classe estimada pelo classificador.

	Combustível'	Restauração'	Saúde'	Supermercado'
Combustível	1			
Restauração		1		
Saúde				1
Supermercado		1		

Tabela 8 – Matriz de Confusão para a 1ª iteração.

	Combustível'	Restauração'	Saúde'	Supermercado'
Combustível				1
Restauração		1		
Saúde			1	
Supermercado		1		

Tabela 9 – Matriz de Confusão para a 2ª iteração.

	Combustível'	Restauração'	Saúde'	Supermercado'
Combustível	1			
Restauração		1		
Saúde			1	
Supermercado		1		

Tabela 10 – Matriz de Confusão para a 3ª iteração.

	Combustível'	Restauração'	Saúde'	Supermercado'
Combustível	1			
Restauração		1		
Saúde			1	
Supermercado				1

Tabela 11 – Matriz de Confusão para a 4ª iteração.

	Combustível'	Restauração'	Saúde'	Supermercado'
Combustível	1			
Restauração		1		
Saúde			1	
Supermercado				1

Tabela 12 – Matriz de Confusão para a 5ª iteração.

Nos testes realizados, a probabilidade de erro anulou-se ao fim de 5 iterações (ver Figura 34), totalizando 24 recibos (6 recibos por cada uma das 4 classes). É preciso referir que, uma das razões para se ter conseguido atingir 100% de taxa de sucesso em poucas iterações, foi o facto de os recibos da mesma classe serem muito semelhantes entre si.

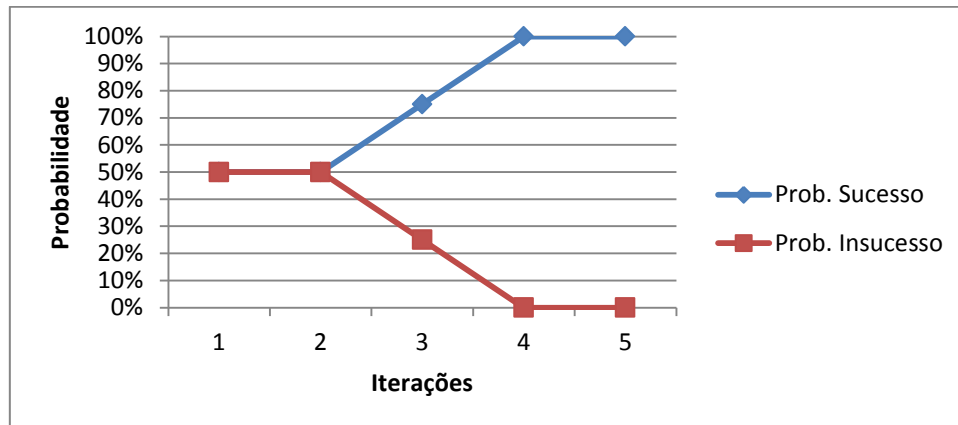


Figura 34 – Probabilidade de sucesso e insucesso do classificador ao longo das iterações.

4.7 Interface Gráfica

Como foi referido anteriormente, a aplicação desenvolvida neste trabalho dispõe de uma interface gráfica com realidade aumentada, para auxiliar o utilizador no processo de detecção do recibo e também de visualização de informação.

Um dos processos que utiliza realidade aumentada é a detecção da posição e dos contornos do recibo em tempo real. Na Tabela 13 são apresentadas as durações médias de cada uma das operações necessárias para fazer a detecção da posição e dos contornos do recibo.

Processo	Duração média (ms)
Escalamento para dimensão inferior	30
Filtragem com filtro de média	54
Binarização	14
Conversão para níveis de cinzento	13
Detecção de contornos de Canny	54
Transformada de Hough e filtragem de rectas	154
Total	319

Tabela 13 – Duração média dos processos envolvidos na detecção com realidade aumentada.

Com base nos valores presentes na Tabela 13 conclui-se que com um total de 319 ms de duração, é possível processar aproximadamente 3 imagens por segundo para apresentar a sobreposição dos contornos do recibo como exemplificado na Figura 35.



Figura 35 – Exemplo de realidade aumentada para detecção dos contornos do recibo.

Além da sobreposição dos contornos do recibo durante a sua detecção, é ainda realizada sobreposição da informação extraída do recibo e resultante do processamento do mesmo. Esta sobreposição é interactiva uma vez que permite a alteração dos dados por parte do utilizador caso estes estejam incorrectos ou incompletos (ver Figura 36).



Figura 36 – Sobreposição do resultado do processamento do recibo.

Conclusões e Trabalho Futuro

Com o trabalho desenvolvido nesta dissertação, foi possível criar uma aplicação para dispositivos móveis para efectuar a digitalização, detecção e classificação automática de recibos em papel. Esta aplicação consegue processar imagens de modo a obter uma imagem normalizada do recibo sem efeitos de perspectiva ou rotação, consegue extrair o texto do recibo e os valores mais relevantes (total e número de identificação fiscal), e classificar o recibo numa determinada classe. Adicionalmente, a aplicação utiliza ainda realidade aumentada para auxiliar o utilizador no processamento de recibos. Assim, conclui-se que os objectivos propostos inicialmente foram alcançados.

Durante a realização do trabalho foram encontradas algumas dificuldades que tornaram este trabalho mais rico pois, foi possível verificar algumas das limitações que ainda existem quando se desenvolve uma aplicação deste género para dispositivos móveis. Estas limitações fizeram com que fosse necessário voltar atrás em algumas das decisões e convicções iniciais, para procurar e testar alternativas, de forma a resolver essas mesmas limitações.

Alguns destes problemas poderão estar resolvidos dentro de algum tempo com a evolução natural da tecnologia e de todos os ecossistemas envolvidos. Destacam-se assim três componentes base que deram origem aos problemas encontrados:

1. Evolução da largura de banda – com o aparecimento das redes móveis 4G e melhoria das restantes redes, o tempo de

transferência de uma imagem para um servidor para efectuar o OCR irá diminuir consideravelmente.

2. Evolução do hardware dos dispositivos móveis – como consequência da Lei de Moore [1], os dispositivos móveis irão continuar a tornar-se cada vez mais avançados em termos de poder de processamento. A ênfase na multiplicação dos núcleos dos processadores em vez da velocidade de relógio irá aumentar o processamento em paralelo que é possível realizar. Os *chips* gráficos também estão a evoluir rapidamente permitindo a execução de aplicações 3D com efeitos complexos, como por exemplo jogos. Poderá ser interessante aproveitar a capacidade de processamento dos GPUs para auxiliar no processamento de informação e execução de algoritmos mais complexos.
3. Evolução das bibliotecas – algumas das bibliotecas que implementam os algoritmos utilizados, nomeadamente o OpenCV e o Tesseract, apesar de terem sido compilados para o hardware do dispositivo móvel, não foram criados de raiz para este tipo de arquitecturas pelo que não estarão optimizados para esta plataforma e aproveitar os recursos disponíveis de forma eficiente.

Posto isto, seria interessante reavaliar, dentro de um prazo que faça sentido, os campos acima referidos para verificar se houve melhorias, nomeadamente, o compromisso entre processamento local ou remoto de operações mais exigentes, como por exemplo o OCR.

Por outro lado, apesar de muitos dos algoritmos utilizados serem computacionalmente exigentes mas dado que existem aplicações no mercado, como por exemplo jogos, que mostram objectos 3D de alta qualidade com texturas e com os mais diversos efeitos em tempo real, acredita-se que seja possível reduzir drasticamente o tempo de processamento responsável pela transformação homográfica. Esta parte será analisada e melhorada em trabalho futuro.

Em termos de OCR, apesar do tempo de processamento ter sido diminuído drasticamente com a passagem do OCR para um servidor remoto,

será necessário afinar melhor o Tesseract, alterando as suas configurações, para que o texto detectado tenha melhor qualidade.

Para concluir será também realizado, em trabalho futuro, um estudo de usabilidade da interface desenvolvida.

Bibliografia

Apêndice

A.1 Diagrama de Classes

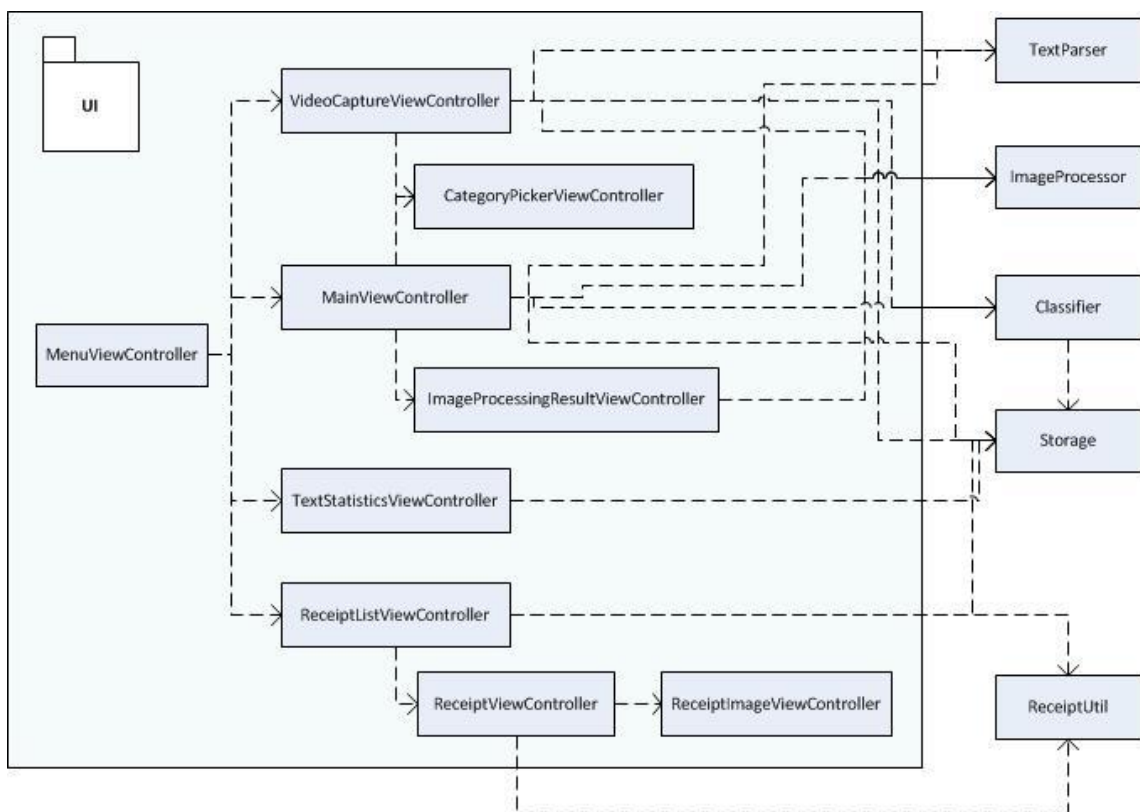


Figura 37 – Diagrama de classes da aplicação.

A.2 Interface Gráfica e Utilização

Nesta secção é realizada uma breve descrição da interface gráfica, explicando como se utiliza a aplicação. As imagens da interface contêm

numeração referente a pontos de interação relevantes que são referidos na descrição de cada imagem.

Na Figura 38 é apresentado o menu inicial da aplicação, tendo como opções disponíveis o processamento de recibos em tempo real (1), o processamento de recibos em imagens gravadas previamente (2), a consulta da lista de recibos já classificados (3) e a consulta da lista de termos mais comuns em cada classe de recibo (4).



Figura 38 – Menu inicial.

A lista de recibos já classificados permite ao utilizador consultar todos os recibos classificados anteriormente e editar os dados caso seja necessário (ver Figura 39). A lista contém um resumo dos dados que poderão ser melhor visualizados quando o utilizador escolhe um dos registos (2). O utilizador pode voltar ao menu inicial utilizando o botão (1).



Figura 39 – Lista de recibos já classificados.

A Figura 40 mostra um exemplo de ecrã de detalhes de um recibo já classificado. Em (2) o utilizador pode activar o modo de edição dos dados do recibo. Caso o utilizador carregue em (3), é mostrada uma versão em alta resolução da imagem do recibo em modo de ecrã cheio. Carregando em (1), o utilizador volta à lista de recibos já classificados.

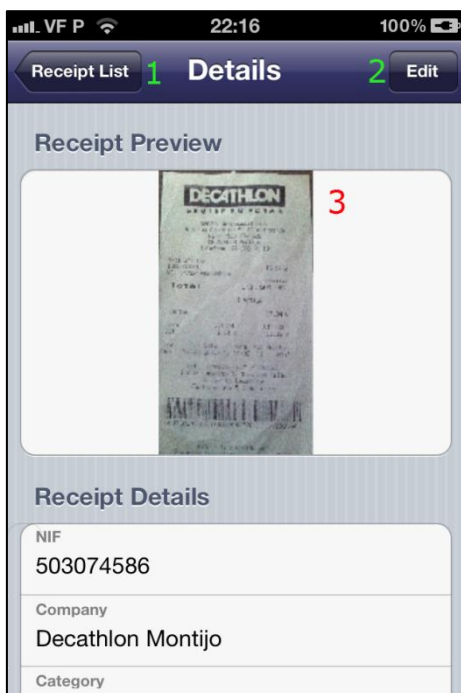


Figura 40 – Detalhes de um recibo já classificado.

O ecrã de apresentação dos termos mais comuns mostra ao utilizador uma lista com os 5 termos mais comuns em cada uma das classes, caso existam dados relativos a essa classe (ver Figura 41). Em (1) o utilizador pode voltar ao menu inicial.



Figura 41 – Lista de termos mais prováveis por classe.

Na Figura 42 é apresentada a interface para processamento de imagens pré-gravadas. Em (4) o utilizador acede a um menu onde pode carregar uma imagem (ver Figura 43). Após o carregamento da imagem, o fundo preto é substituído pela imagem carregada. Em (2) é realizado o processo de detecção do recibo, sendo a imagem carregada do recibo substituída pela versão normalizada da imagem (contendo apenas o recibo centrado). Em (3) é realizado o OCR sobre a imagem normalizada e, após a conclusão do OCR, é realizada a classificação do recibo. O resultado da classificação é apresentado sobreposto à imagem do recibo, como mostra a Figura 45. Em (1) o utilizador pode voltar ao menu inicial.

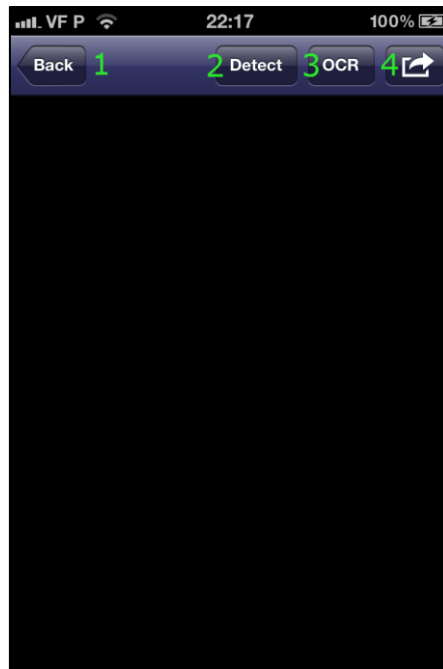


Figura 42 – Processamento de fotografias antes de escolher uma imagem de um recibo.

Como já referido anteriormente, a Figura 43 mostra o menu apresentado ao utilizador para carregar a imagem a ser processada. Este menu permite carregar uma imagem da galeria de fotografias do dispositivo móvel (1) ou carregar uma das imagens de recibos de teste incluídas na aplicação (2). Caso já tenha sido carregada uma imagem, em (3) é possível iniciar o modo passo-a-passo onde o utilizador pode ver todos os processos aplicados à imagem do recibo.

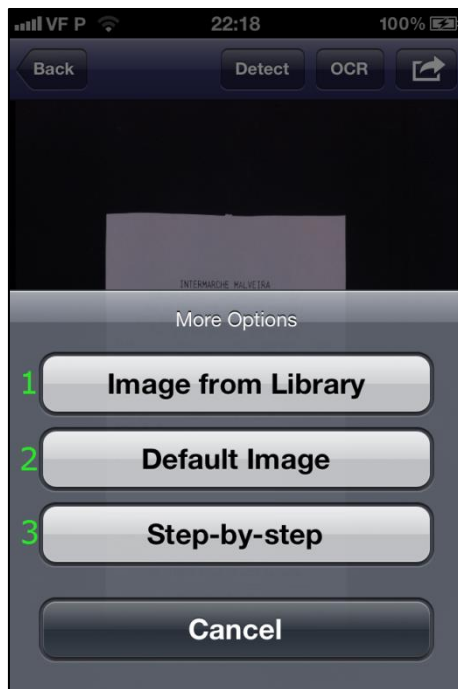


Figura 43 – Menu para carregar imagens de recibos ou iniciar modo passo-a-passo.

No modo passo-a-passo o utilizador pode ver o resultado de cada operação realizada sobre a imagem e o tempo que demorou a realizar essa operação. Em (1) o utilizador volta ao ecrã de processamento de imagens, mostrado na Figura 44. Em (2) o utilizador dá início à operação seguinte. Quando é realizada a última operação, a única opção será voltar para trás carregando em (1).

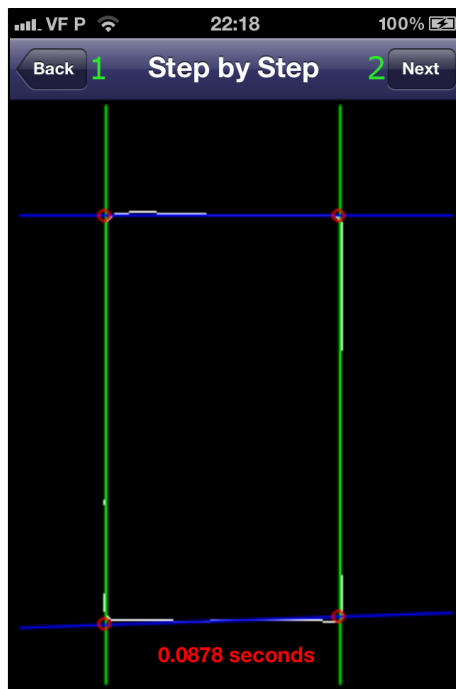


Figura 44 – Modo passo-a-passo.

A apresentação dos resultados do processamento do recibo é feita através da sobreposição da informação à imagem do recibo (ver Figura 45). Aqui o utilizador pode editar os valores caso estes estejam incorrectos. Em (1) é possível editar o número de identificação fiscal (NIF). Em (2) é possível adicionar o nome da empresa, que após ser adicionado irá aparecer automaticamente sempre que o mesmo NIF for detectado. Em (3) é possível editar a classe do recibo e finalmente em (4), é possível editar o valor do total do recibo.



Figura 45 – Processamento e classificação de imagem terminados.

Na Figura 46 é apresentado o início do modo de vídeo, em que o utilizador aponta a câmara do dispositivo móvel para o recibo e tenta enquadrá-lo na imagem. Este processo recorre a realidade aumentada para auxiliar o utilizador no enquadramento do recibo, uma vez que este consegue visualizar os contornos detectados em tempo real. Quando o utilizador estiver satisfeito com o enquadramento do recibo na imagem pode dar início ao processo em (2). Se o utilizador desejar voltar ao menu inicial, poderá fazê-lo carregando em (1).



Figura 46 – Modo vídeo: detecção de contornos em tempo real com realidade aumentada.

Concluído o processamento do recibo, é apresentado o resultado sobre o vídeo em tempo real (ver Figura 47). Caso o NIF e o total não tenham sido extraídos com sucesso, é apresentada uma mensagem ao utilizador para este considerar repetir o processo (6) uma vez que, por falta de qualidade do texto detectado, não foi possível extrair os valores. Aqui, à semelhança do que acontece na Figura 45, o utilizador poderá editar o resultado, caso encontre erros. Em (7) o utilizador pode gravar os dados e em (1) pode voltar ao menu inicial.



Figura 47 – Modo vídeo: Processamento e classificação do recibo terminados.