Universidade Federal de Pernambuco
Graduação em Ciência da Computação
Centro de Informática

FERRAMENTA MOBILE PARA RECONHECIMENTO VISUAL DE FONTES

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Pedro Augusto Silva Lucena (pasl@cin.ufpe.br)
Orientador: Cleber Zanchettin (cz@cin.ufpe.br)

Recife, 14 de Setembro de 2016

1.Contexto

A tipografia é o principal elemento de design de qualquer texto. Designers gráficos possuem interesse em fontes, tanto as de seus trabalhos como de outros Designers. Consequentemente, eles com frequência se deparam com o problema de identificação dos tipos de fontes utilizadas em diferentes trabalhos no mundo real [1], como por exemplo a fonte utilizada em uma propaganda ou em um menu de restaurante.

O Reconhecimento Visual de Fontes (VFR) é tratado pela área da computação chamada visão computacional, em conjunto com sub-áreas tais como aprendizagem de máquina. As fases da identificação incluem o reconhecimento de caracteres na imagem, a separação dos caracteres e o teste destes em uma rede neural pré-treinada com uma certa quantidade de fontes. Um importante conceito que auxilia a identificação de fontes é o de Reconhecimento Óptico de Caracteres (OCR), que já é bem difundido e amplamente estudado pela comunidade científica e possui ferramentas prontas disponíveis para uso.

Atualmente existem algumas ferramentas que se propõem a oferecer esta identificação, porem a grande maioria delas funciona em plataformas web, ou seja, não resolve o problema de maneira imediata, visto que o usuário precisaria tirar uma foto em determinado momento e então ao chegar em um lugar com computador e disponibilidade, transferir esta foto para o site e aguardar o resultado. O WhatTheFont Mobile [2] até o momento é a única ferramenta em formato de aplicativo para smartphones, porém requer uma quantidade alta de entradas por parte do usuário para permitir a análise da fonte.

Investigando o universo de aplicações voltadas para a identificação de tipografia por imagem, ficou claro que há uma defasagem na questão comercial. Na pesquisa acadêmica existem diversos artigos oferecendo novas soluções, menos custosas e consequentemente mais rápidas, porém, não aplicadas além dos casos de teste. Entre essas duas visões surgiu a proposta deste trabalho de graduação em oferecer uma ferramenta com nível comercial para um usuário final com funcionalidades mais refinadas a partir de ideias discutidas em artigos acadêmicos.

2.Proposta

A proposta deste trabalho de graduação é desenvolver uma ferramenta para dispositivos móveis que reconheça a fonte de um determinado texto a partir da captura de uma imagem do texto pela câmera do dispositivo. A ideia é oferecer maior robustez e o mínimo de intervenção manual por parte do usuário, quando comparado às soluções já existentes.

A ferramenta *open source* Tesseract [3] será utilizada para fazer o reconhecimento de caracteres (OCR). A adição deste reconhecimento deve ser suficiente para automatizar ao máximo o processo, mas existem ainda outras opções que podem ser consideradas. Para isto serão estudados os conceitos de decomposição de caracteres em *Quad-trees* [4] para facilitar reconhecimento; bem como métodos de otimização de busca em bases de dados para reconhecimento de fontes em larga escala como visto em [1]; além de abordagens alternativas como análise de textura global [5] ao invés de reconhecimento de caracteres. Uma combinação das ideias mais viáveis será colocada em prática.

3.Metodologia

Nesta seção apresentamos um resumo da metodologia a ser utilizada ao longo do desenvolvimento do trabalho.

Como já relatado na seção anterior, existe a intenção de incorporar novas abordagens com objetivo de melhorar desempenho e usabilidade, mas em termos de fluxo de desenvolvimento inicialmente tomamos como base um algoritmo padrão de reconhecimento de fontes utilizando o OCR [6].

O primeiro passo é garantir na ferramenta que a entrada do usuário esteja de acordo com um padrão a ser determinado, como por exemplo um número de caracteres definido e angulação do texto. Isto pode ser alcançado restringindo a área de captura da câmera a uma *bounding box.*

Em seguida a captura passará por tratamento para destaque das cores do texto em relação ao background e então reconhecimento por parte do Tesseract. O algoritmo funciona da seguinte maneira: primeiramente é feita a extração dos contornos, estes contornos são agrupados em estruturas chamadas *blobs*, *blobs* por sua vez auxiliam a identificação de linhas de texto e essas linhas são quebradas em palavras a partir do reconhecimento de espaçamento entre caracteres. Cada palavra então é passada por um classificador adaptativo e todas as que são satisfatoriamente reconhecidas são passadas imediatamente como dados de treinamento. Ao fim do texto, todas as palavras são novamente avaliadas para garantir o reconhecimento mais preciso uma vez que o próprio classificador recebeu novas entradas de treinamento [3].

Alternativamente, o reconhecimento pode ser alcançado numa leitura individual de caracteres, dividindo-o em quatro blocos de tamanhos variados baseados no centro de gravidade da imagem. Este método é chamado decomposição em *quad-trees* [4].

Uma vez tendo os caracteres reconhecidos salvos separadamente, cada um deles será testado para classificação a partir de uma rede neural treinada com uma quantidade de fontes distintas, espera-se um resultado comum das várias entradas e este resultado da classificação então é retornado para o usuário como sendo a melhor aproximação da fonte do texto de entrada. No caso da decomposição em quad-trees, as partes dos caracteres são avaliadas em uma árvore métrica contendo imagens de caracteres de um banco de dados de fontes. Nesse caso, o resultado é mais absoluto e menos aproximado, porém tende a ser mais custoso quanto maior a quantidade de fontes no banco.

Outra opção que pode ser considerada é a de analisar a textura global do texto e oferecer o reconhecimento da fonte. Para isto, a entrada do usuário deve conter um pequeno texto, e não só uma palavra ou frase. Após o pré processamento para retirada de espaçamentos desnecessários é feita a extração de características a partir da aplicação do filtro de Gabor. Em seguida um algoritmo de reconhecimento de padrões classifica essas características de acordo com as fontes treinadas.

A princípio, esta ferramenta será desenvolvida para o sistema operacional iOS, utilizando a linguagem de programação Swift e o ambiente de desenvolvimento Xcode.

4.Cronograma

O cronograma exposto na tabela abaixo descreve as atividades deste trabalho de graduação com divisão dos meses em semanas:

	Mês															
Atividade	**Setembro**				**Outubro**				**Novembro**				**Dezembro**			
Revisão bibliográfica	■	■	■	■												
Criação da base de dados			■	■	■	■										
Desenvolvimento do aplicativo					■	■	■	■	■	■						
Análise dos resultados									■	■	■	■				
Relatório do trabalho							■	■	■	■	■	■	■	■		
Defesa															■	

5.Referências Bibliográficas

[1] Chen, G., Yang, J., Jin, H., Brandt, J., Shechtman, E., Agarwala, A. and Han, T.X. (2014) 'Large-scale visual font recognition', *2014 IEEE Conference on Computer Vision and Pattern Recognition*, . doi: 10.1109/cvpr.2014.460.

[2] Inc2016MyFonts (1999) *WhatTheFont mobile: Overview « MyFonts*. Available at: https://www.myfonts.com/WhatTheFont/mobile/ (Acessado: 8 de setembro de 2016).

[3] GitHub (2016) *Tesseract-ocr*. Available at: https://github.com/tesseract-ocr (Acessado: 14 de setembro de 2016).

[4] Sexton, A., Todman, A. and Woodward, K. (no date) *Font recognition using shape-based Quad-tree and Kd-tree decomposition*. Available at: http://www.cs.bham.ac.uk/~aps/research/papers/pdf/SeToWo_CVPRIP00-FontRecognitionUsingShape-BasedQuad-TreeAndKD-TreeDecomposition.pdf (Acessado: 8 de setembro de 2016).

[5] Zhu, Y., Tan, T. and Wang, Y. (1999) 'Font recognition based on global texture analysis', Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR '99 (Cat. No.PR00318), . doi: 10.1109/icdar.1999.791796.

[6] Khoubyari, S. and Hull, J.J. (1996) 'Font and function word identification in document recognition', *Computer Vision and Image Understanding*, 63(1), pp. 66–74. doi: 10.1006/cviu.1996.0005.

6.Assinaturas

Prof. Cleber Zanchettin
Orientador

Pedro Augusto Silva Lucena
Aluno