DETECÇÃO E RECONHECIMENTO DE GESTOS PARA INTERAÇÃO HUMANO-COMPUTADOR

GESTURE DETECTION AND RECOGNITION FOR HUMAN-COMPUTER INTERACTION

Wellington Pires da Silva [1]
Osvandre Alves Martins [2]

Data de entrega dos originais à redação em: 29/04/2017
e recebido para diagramação em: 19/07/2018

A evolução tecnológica tem proporcionado novas formas de interação das pessoas com máquinas e equipamentos, principalmente os computadores. Nesse contexto, determinados sensores integrados a dispositivos de tamanho reduzido e com poder de processamento de dados representam um meio de se capturar e traduzir movimentos realizados por uma pessoa em comandos de operação de sistemas computadorizados, tornando esta ação mais natural e intuitiva. Apresentam-se os detalhes de uma abordagem para a interação humano-computador baseada em gestos e no emprego integrado de acelerômetro, de tecnologias de processamento e comunicação de dados e na aplicação conjunta das técnicas de reconhecimento de padrões em séries temporais DTW (Dynamic Time Warping) e KNN (K-Nearest Neighbors) associadas a um vocabulário de gestos mapeados em comandos de operação de um aplicativo de software. Por meio de protótipos na forma de tecnologia vestível, foi possível verificar e validar a abordagem proposta operando software aplicativo instalado em um computador pessoal convencional.

Palavras-chave: Interação Humano-Computador. Detecção e Reconhecimento de Gestos. Reconhecimento de Padrões. Tecnologia Vestível.

The technological evolution has been providing new ways for people to interact with machines and equipment, more specifically with computers. In this context, specific sensors integrated to small devices able to process data represent a way of getting and translate data about movements performed by a person into computerized system operation commands, making this action more natural and intuitive. This paper presents details about an approach for human-computer interaction based on gesture, as well as the integrated use of accelerometer, data processing and communication technologies, and the joint application of the time series pattern recognition techniques DTW (Dynamic Time Warping) e KNN (K-Nearest Neighbors) associated to gesture vocabulary mapped into software application commands. By means of prototypes as wearable technology, it was possible to verify and validate the proposed approach, by operating a software application installed on a standard personal computer.

Keywords: Human-Computer Interaction. Gesture Detection and Recognition. Patterns Recognition. Wearable Technology.

1 INTRODUÇÃO

O reconhecimento de gestos realizados com as mãos mostra-se como uma alternativa atraente e natural para a Interação Homem-Máquina, principalmente a Interação Humano-Computador (IHC), se comparado a outras formas empregadas tradicionalmente e baseadas no acionamento de dispositivos como teclados e mouses. Isso ocorre pois, em determinadas situações, tais dispositivos ainda se mostram incompatíveis para IHC. A título de exemplo, cita-se a interação com objetos 3D em determinados aplicativos de software, pois o mouse não consegue emular corretamente as suas 3 dimensões (RAUTARAY; AGRAWAL, 2015).

Entre as diferentes abordagens para um sistema de computação reconhecer gestos realizados com as mãos, encontram-se aquelas que consideram o emprego de sensores na forma de acelerômetro. Segundo Mace, Gao e Coskun (2013), esta abordagem é uma importante área de interesse para IHC, apresentando resultados úteis para evoluções em seu contexto.

Este trabalho apresenta os detalhes da concepção de uma solução tecnológica envolvendo a aplicação de *hardware* e de *software* para possibilitar o treinamento, a detecção e o reconhecimento de gestos realizados com a mão, mapeando esses gestos em comandos de operação de um sistema computadorizado. Em termos de algoritmos, consideram-se como bases o reconhecimento de padrões, o pré-processamento e a segmentação de dados.

2 RECONHECIMENTO DE PADRÕES

Segundo Kpalma e Ronsin (2007), o reconhecimento de padrões é uma disciplina científica da área de Inteligência Artificial, mais especificamente da Aprendizagem de Máquina que entre outras técnicas considera a classificação de dados (padrões) em uma série de categorias ou classes. Exemplos de padrões podem ser um pixel em uma imagem ou forma 2D ou 3D, gesto, impressão digital, rosto humano, e voz, representados em séries temporais (conjunto de valores medidos ao longo de um determinado período de tempo).

Ainda segundo Kpalma e Ronsin (2007), sistemas de reconhecimento de padrões visam identificar classes

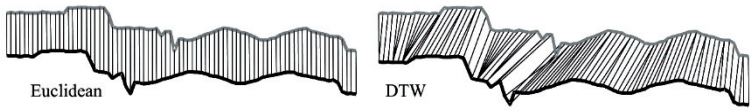
1 - Tecnologia em Análise e Desenvolvimento de Sistemas - Instituto Federal São Paulo (IFSP) - Câmpus Votuporanga. <wellington.pires@live.com >.
2 - Instituto Federal São Paulo (IFSP) - Câmpus Votuporanga. < osvandre@ifsp.edu.br >.

de valores, tentando obter, de forma automática, as mais prováveis. Dentre os algoritmos com possível aplicação para tanto, se encontram o *Dynamic Time Warping* (DTW) e o *K-Nearest Neighbours* (KNN), ambos apresentados a seguir.

2.1 Dynamic Time Warping (DTW)

O algoritmo DTW realiza o cálculo da similaridade entre duas series temporais de dimensões diferentes. Keogh e Ratanamahatana (2005) explicam o DTW com base na Figura 1 que ilustra dois alinhamentos de series temporais, um utilizando a Distância Euclidiana (*Euclidean*) e outro utilizando o DTW. Percebe-se que, embora as duas series tenham uma forma global similar, elas não estão alinhadas em relação ao eixo x. A Distância Euclidiana assume que o i-ésimo ponto em uma série-base é alinhado com o i-ésimo ponto de outra série, produzindo uma medida de similaridade pessimista. O DTW por sua vez, possibilita uma medida de similaridade mais intuitiva, combinando cálculo de Distância Euclidiana com Programação Dinâmica.

Figura 1 - Comparação de series temporais utilizando Distância Euclidiana e DTW



Fonte: Keogh e Ratanamahatana, 2005

2.2 K-Nearest Neighbors (KNN)

Segundo Yu et al. (2011, p. 282), o algoritmo KNN é capaz de classificar objetos com base na sua proximidade com objetos de treino vizinhos em um espaço de características. Trata-se de uma espécie de aprendizagem na forma *lazy learning* (aprendizagem preguiçosa) que determina a proximidade entre um dado e outro e posterga a realização de uma série de cálculos para o momento da classificação. Esta classificação, ainda segundo Yu et al. (2011, p. 282), é realizada por meio do voto da maioria dos k vizinhos mais próximos de cada dado, sendo k um número inteiro positivo. Os vizinhos formam um conjunto de objetos de treino, cuja classificação correta é previamente conhecida. A sua definição se baseia em vetores de posição no espaço de características multidimensionais e a Distância Euclidiana ou outras formas de medida como a Manhattan, por exemplo, podem ser empregadas para se determinar a proximidade da vizinhança.

3 PRÉ-PROCESSAMENTO E SEGMENTAÇÃO

O conjunto de dados de aceleração obtidos a partir de acelerômetro normalmente apresenta valores acima ou abaixo da média das leituras realizadas. Estes valores são conhecidos como ruídos e necessitam ser desconsiderados, caracterizando uma atividade de pré-processamento. Para tanto, empregou-se o filtro de passa-baixas, descrito pela equação I obtida a partir do trabalho de Mercer (2003). Este realiza a "suavização" dos valores obtidos, viabilizando uma segmentação mais precisa.

$$y_t = ax_t + (1-a)y_{t-1} \qquad (1)$$

A segmentação se refere a um passo importante para o mapeamento de um gesto, tendo como base uma série temporal obtida a partir de leituras pelo acelerômetro, realizadas de tempo em tempo. Desta forma, é possível identificar quando um gesto foi iniciado e quando ele foi finalizado.

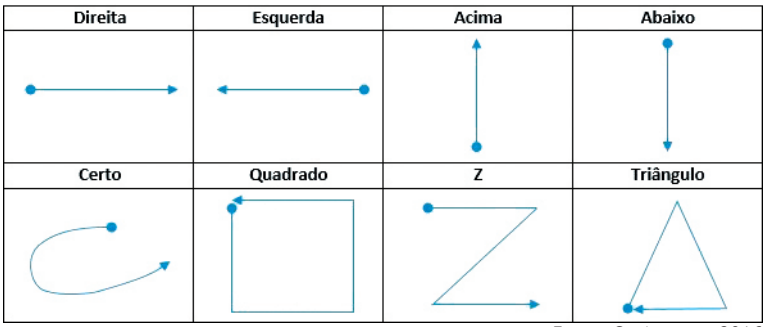
Neste trabalho o processo de segmentação se baseia na definição apresentada no trabalho de Prekopcsák (2008) e que afirma que "um gesto inicia com uma rápida aceleração, muda de direção continuamente, termina em posição quase constante e dura mais que 0,8 segundos".

4 TREINO E RECONHECIMENTO DE GESTOS

O reconhecimento de padrões é aplicado, com a estratégia supervisionada, sobre o conjunto de séries temporais que representam gestos. Estas séries são consideradas treinos e são obtidas por meio da execução de gestos pré-definidos em um vocabulário. A Figura 2 apresenta uma possível configuração de vocabulário de forma que o pequeno círculo indica o início do gesto e a ponta da seta o final deste.

Uma vez registrados os treinos para o vocabulário, as técnicas KNN e DTW podem ser empregadas conjuntamente, para realizar o reconhecimento de gestos.

Figura 2 – Exemplo de configuração de vocabulário de gestos

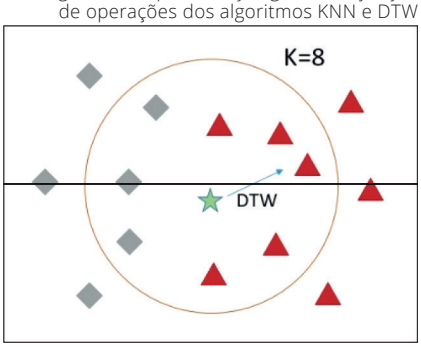


Fonte: Os Autores, 2016

A Figura 3 ilustra esta abordagem e estratégia constatada a partir do trabalho de Regan (2014), espelhado nos relatos de Mitsa (2010). Percebe-se que o gesto a ser reconhecido (estrela) tem a sua série temporal comparada por meio do DTW com todos os outros gestos presentes na base de treinamento (losangos e triângulos).

Realizada a comparação, escolhem-se k vizinhos mais próximos, de forma que k foi definido, empiricamente como 8 (oito), mediante observações em testes do algoritmo. Estes k vizinhos (circulo) são utilizados para fazer uma votação do gesto que aparece mais vezes, a partir do resultado classifica-se o gesto a ser reconhecido.

Figura 3 – Representação gráfica da junção de operações dos algoritmos KNN e DTW



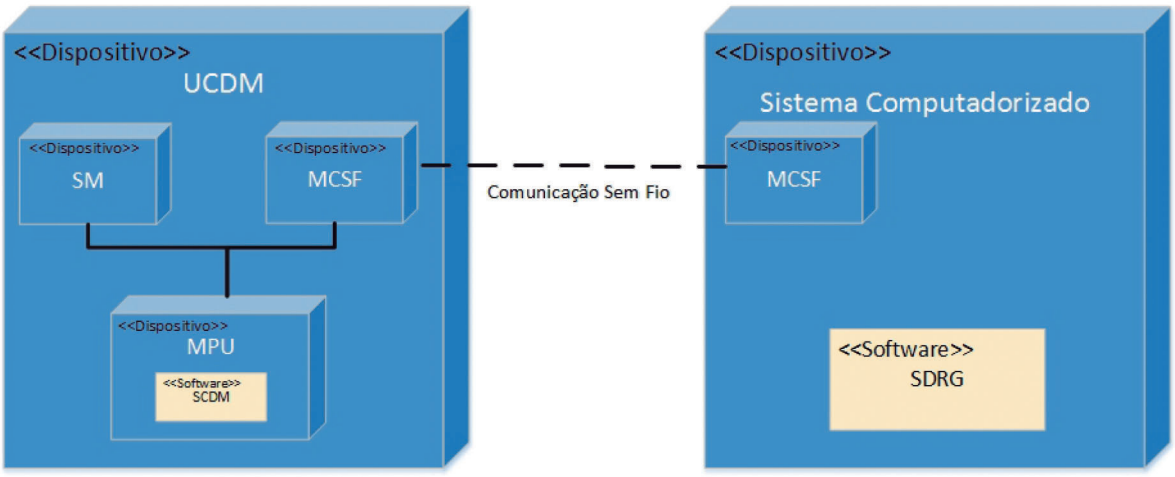
Fonte: Os Autores, 2016

Devido a questões de eficiência e precisão, constatou-se a necessidade de modificar o algoritmo DTW padrão, de modo a considerar 3 séries temporais ao invés de uma. Esta modificação se refere à utilização de cada valor de aceleração dos eixos x, y e z em separado, ao invés de fazer algum tipo de tratamento para se obter uma série temporal referente à média deles, por exemplo.

5 PROTOTIPAGENS

Embora o foco deste trabalho se encontra em *software* que possibilite o reconhecimento de gestos, seguido de sua tradução para comandos de operação de um sistema computadorizado, as implementações de soluções para interação humano-computador geralmente consideram aspectos de *hardware* também. Para testes da abordagem apresentada, elaborou-se um protótipo cuja visão estrutural de sua arquitetura é ilustrada na Figura 4.

Figura 4 - Visão estrutural da arquitetura de sistema do protótipo desenvolvido para testes da solução proposta



Fonte: Os Autores, 2016

Notem-se os seguintes componentes: **UCDM (Unidade de Coleta de Dados de Movimentos) -** dispositivo que pode ser vestido pelo usuário e que abriga o *hardware* necessário para leitura de dados de sensores, bem como para a sua comunicação ao *software* voltado à detecção e identificação dos movimentos; **SM (Sensor de Movimento) –** acelerômetro integrado à UCDM, responsável por coletar dados de aceleração referentes aos movimentos efetuados pelo usuário; **MCSF (Módulo de Comunicação de Dados Sem Fio) -** componente da UCDM e do sistema computadorizado, responsável por constituir uma rede de comunicação de dados baseada no padrão IEEE 802.15 (*Bluetooth*); **MPU (Microcontrolador de Placa Única) –** *hardware* principal da UCDM com *software* embarcado, responsável por integrar a SM e o MCSF; **SCDM (*Software* de Coleta de Dados de Movimentos)** - *software* embarcado na MPU, responsável pela coleta dos dados do SM e envio por meio do MCSF; **SDRG (*Software* de Detecção e Reconhecimento de Gestos)** - *software* instalado em sistema computadorizado, responsável pelos cálculos necessários à detecção e identificação dos gestos; e **Sistema Computadorizado** - microcomputador dotado de sistema operacional que suporte a execução do SDRG e que precise ser operado pelo usuário para alguma finalidade.

5.1 Implementações em hardware

Como resultado de um desenvolvimento iterativo e incremental, duas versões diferentes de protótipo para os testes foram implementadas. A primeira versão, ilustrada na Figura 5A, considera energização por meio de cabo conectado à porta USB de microcomputador e os seguintes componentes de *hardware* foram empregados: **Arduino Pro Mini -** plataforma de prototipagem eletrônica dotada de microcontrolador, implementando a MPU da arquitetura (A); **MPU6050 -** sensor de alta tecnologia em processamento de movimento, combinando acelerômetro e giroscópio, implementando o SM da arquitetura (B); e **HC-05 -** módulo *Bluetooth*, que possibilita o envio e recepção de dados, implementado o MCSF da arquitetura (C).

No sentido de remover a energização por cabos, buscou-se por placas de circuito que utilizassem baterias como fonte de energia. Deparou-se então com a *Realtag iBeacon*, da empresa chinesa *Bytereal*, energizada por meio de uma bateria de Lítio CR2032 e que já possui um *Bluetooth Low Energy* (CC2541) e um acelerômetro (MPU6050) integrados. Também possui *software* embarcado próprio que realiza leituras e comunicações de dados do acelerômetro em tempo programado. Desta forma, constatou-se que ele é capaz de implementar toda a UCDM, conforme ilustrado na Figura 5B.

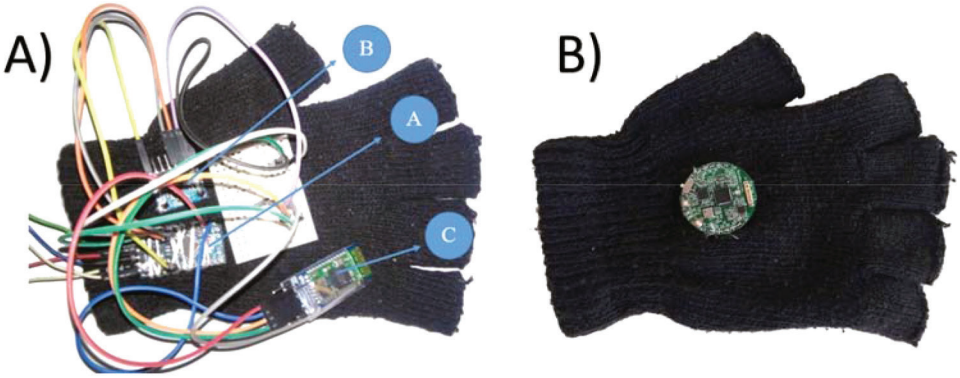
5.2 Implementações em software

As implementações em *software* se referem aos componentes SCDM e SDRG da arquitetura apresentada anteriormente. A primeira versão do SCDM se refere a um código de programação para Arduino, utilizado na primeira versão de *hardware* do protótipo. Já na segunda versão, o software embarcado da *Realtag iBeacon* desempenha esta função.

Quanto ao SDRG, implementou-se um aplicativo na linguagem C# que recebe os dados coletados pela UCDM a cada 1 (um) segundo e os processa quanto ao reconhecimento do gesto e produção de comando do sistema computadorizado em que se encontra instalado.

A associação de itens do vocabulário de gestos com comandos efetivos de operação do sistema computadorizado foi implementada, provisoriamente, de forma direta no código fonte (*hard coding*). No caso, um

Figura 5 – Protótipos de hardware da tecnologia vestível desenvolvidos para testes



Fonte: Os Autores, 2016

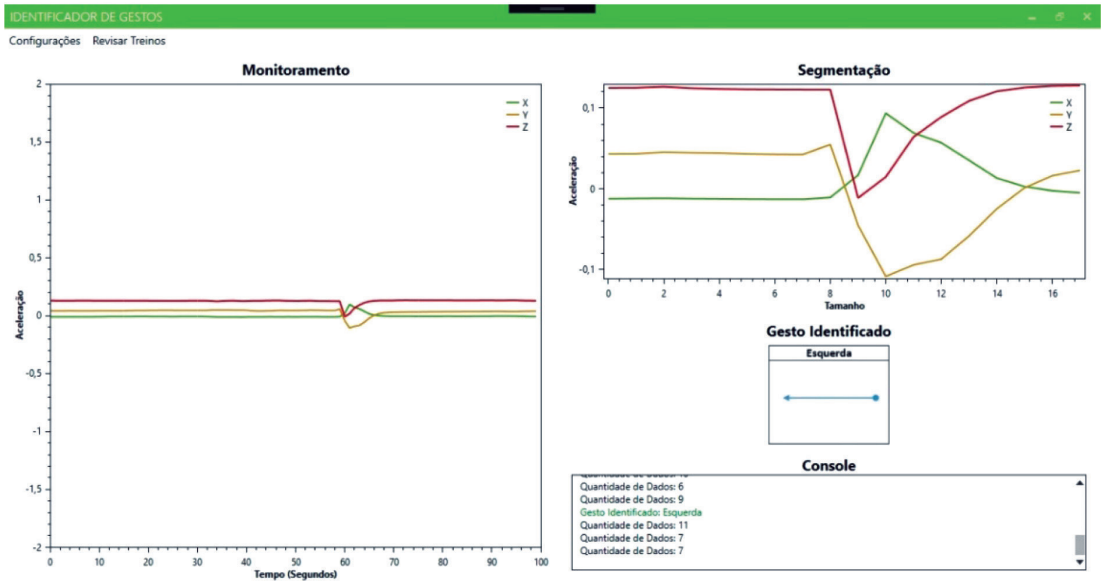
conjunto de gestos referentes ao vocabulário ilustrado na Figura 2 foi selecionado para produzir eventos referentes ao pressionamento de teclas de um teclado (**Direita** = seta à direita, **Esquerda** = seta à esquerda, **Acima** = seta para cima e **Abaixo** = seta para baixo). Os demais gestos do vocabulário podem ser associados a outras teclas, conforme necessidades de aplicação.

A Figura 6 ilustra a interface do protótipo do software desenvolvido para viabilizar treinos, detecção e reconhecimento de gestos. O gráfico de Monitoramento possibilita a visualização de movimentos capturados pelos sensores em um intervalo do tempo. O gráfico de Segmentação representa um detalhamento ou zoom da faixa de valores em que se detectou alteração de padrão dentro deste intervalo de tempo e cuja série temporal se aproxima da série referente a um item do vocabulário de gestos treinados. Uma vez identificado pelo software, uma figura que expressa o referido gesto é apresentada, bem como mensagens em um console voltado a prover mais informações sobre o processamento dos dados pelo algoritmo.

6 RESULTADOS E DISCUSSÕES

Para testes de verificação e validação da solução implementada, constituiu-se uma base de treinos. Cada gesto do vocabulário foi treinado 20 (vinte) vezes, produzindo um conjunto de 240 (duzentos e quarenta) registros. Para cada gesto treinado, foi realizada a sua execução 20 (vinte) vezes, verificando a precisão em seu reconhecimento. Os resultados são apresentados na Tabela 1.

Figura 6 – Interface do protótipo do software de detecção e reconhecimento de gestos



Fonte: Os Autores, 2016

Tabela 1 - Testes de Precisão no Reconhecimento de Gestos

Gesto	Acerto	Erro	Precisão
Esquerda	19	1	95%
Direita	20	0	100%
Acima	19	1	95%
Abaixo	20	0	100%
Certo	19	1	95%
Quadrado	17	3	85%
Z	20	0	100%
Triângulo	18	2	90%
		Média	95%

Tabela 1 - Testes de Precisão no Reconhecimento de Gestos

Observa-se uma precisão média de 95% no reconhecimento de gestos, fato que indica possível viabilidade de aplicação da solução projetada. Salienta-se que para alguns gestos obteve-se precisões menores que 90%. Analisando os casos, constatou-se que isto ocorre devido a alguns gestos possuírem similaridades em sua representação numérica em série temporal, por exemplo, os gestos "Esquerda" e "Quadrado" definidos no vocabulário. Esta ocorrência de baixa precisão pode se agravar à medida que a quantidade de gestos no vocabulário aumentar. Diante disso, acredita-se que a solução proposta possua limitação de aplicação a casos em que a variedade de gestos é elevada.

7 CONCLUSÃO E CONSIDERAÇÕES FINAIS

Este trabalho apresentou conceitos, técnicas e tecnologias inerentes à concepção e elaboração de uma solução tecnológica aplicável na detecção, reconhecimento e mapeamento de gestos em comandos de operação de um sistema computadorizado. Uma precisão de 95% no reconhecimento de gestos foi observada, indicando viabilidade da solução, principalmente para casos em que a quantidade de gestos a reconhecer não for elevada. Constatou-se também que a imprecisão de 5% se refere, principalmente, a gestos com características específicas. Tais limitações representam trabalhos futuros em potencial. Outra necessidade de melhoria se refere a tornar mais fácil a associação de gestos a comandos do sistema computadorizado.

REFERÊNCIAS

KEOGH, E.; RATANAMAHATANA, C. A. Exact indexing of dynamic time warping. **Knowledge and information systems**, v. 7, n. 3, p. 358–386, 2005.

KPALMA, K.; RONSIN, J. An overview of advances of pattern recognition systems in computer vision. **Vision systems: segmentation and pattern recognition**, p. 169–194, 2007.

MACE, D.; GAO, W.; COSKUN, A. Accelerometer-based hand gesture recognition using feature weighted naïve bayesian classifiers and dynamic time warping. **Proceedings of the companion publication of the 2013 international conference on intelligent user interfaces companion - IUI '13 Companion**, p. 83, 2013.

MERCER, C. **Data smoothing: RC filtering And exponential averaging**, 2003. Disponível em: <http://blog.prosig.com/2003/04/28/data-smoothing-rc-filtering-and-exponential-averaging/>. Acesso em: 8 out. 2016

MITSA, T. **Temporal data mining**. 2010.

PREKOPCSÁK, Z. Accelerometer based real-time gesture recognition. **Proceedings of the 12th international student conference on electrical engineering**, 2008.

RAUTARAY, S. S.; AGRAWAL, A. Vision based hand gesture recognition for human computer interaction: a survey. **Artificial intelligence review**, v. 43, n. 1, p. 1–54, 2015.

REGAN, M. **K nearest neighbors & dynamic time warping**, 2014. Disponível em: <https://github.com/markdregan/K-Nearest-Neighbors-with-Dynamic-Time-Warping>. Acesso em: 7 nov. 2016

YU, F. et al. **Three-dimensional model analysis and processing**. 2011.