TEC2014-53176-R HAVideo (2015-2017)

High Availability Video Analysis for People Behaviour Understanding

D2.4v2

Exploration and viability studies for people behaviour understanding

Video Processing and Understanding Lab

Escuela Politécnica Superior

Universidad Autónoma de Madrid

Supported by

AUTHORS LIST

Rafael Martín Nieto	rafael.martinn@uam.es
Marcos Escudero Viñolo	marcos.escudero@uam.es

HISTORY

Version	Date	Editor	Description
0.9	10 December 2015	Rafael Martín	Final Working Draft
1.0	12 December 2015	José M. Martínez	Editorial checking
1.1	13 December 2017	Marcos Escudero Viñolo	Contributions
1.2	16 December 2017	Rafael Martín Nieto	Final Working Draft v2
2.0	18 December 2017	José M. Martínez	Editorial checking

CONTENTS:

1. INTRODUCTION 1

1.1. DOCUMENT STRUCTURE 1

2. HEART RATE DETECTION USING VIDEO 2

3. PEOPLE PRIVACY PRESERVATION IN VIDEO-SURVEILLANCE 5

4. MATERIAL IDENTIFICATION THROUGH THE KINECT TECHNOLOGY 7

5. DESIGN OF SUPPORT TOOLS FOR THE DETECTION OF LOGOS IN VIDEO SEQUENCES 9

6. MATERIAL IDENTIFICATION THROUGH THE KINECT TECHNOLOGY (CONTINUATION) 11

7. APPLICATION OF MULTI-PLANE AUGMENTED REALITY USING CORRESPONDENCES BETWEEN POINTS OF INTEREST 13

8. CONCLUSIONS 15

REFERENCES 17

1. Introduction

This deliverable describes the work related with the task T2.4, exploration and viability studies. Five main works has been developed and are presented below.

The objective of this document is to explore and perform viability studies of technologies that may contribute additionally to the project technologies. The goal of this task is to explore research alternatives not directly considered within the project. It may comprise the use of additional sensors, processing methodologies or the study of new application areas where the developed approaches may be of use (technological re-usability).

All these works have been developed and implemented by the Video Processing and Understanding Lab in the Escuela Politécnica Superior of the Universidad Autónoma de Madrid.

1.1. Document structure

This document contains the following chapters:

- Chapter 1: Introduction to this document
- Chapter 2: Describes a system developed for heart rate detection using video.
- Chapter 3: Describes a work focused on people privacy preservation in video-surveillance.
- Chapter 4: Presents a work which combines colour, texture and depth information to identify materials in a Kinect-like controlled scenario.
- Chapter 5: describes an automatic method for the detection of brand logos in sport sequences.
- Chapter 6: Continues the work presented in chapter 4 considering more advanced techniques and a more elaborated processing.
- Chapter 7: Creates a naïve prototype of augmented reality.
- Chapter 8: Finishes the document with conclusions and future work.

2. Heart rate detection using video

The objective of this work is to design and develop an algorithm that allows to detect the heart rate by analysing natural colour, segmentation masks and depth video sequences [1]. Using this last kind of sequences allows the preservation of the privacy of the monitored person and working in conditions of low or no lighting.

A new approach has been designed and implemented, starting from a base algorithm based on previous work in which we rely. After implementing and validating the base algorithm on colour natural sequences, the new approach has been designed and developed: an algorithm working on colour sequences, as the base algorithm, but also on masks segmentation and depth sequences provided by the Kinect camera.

We have analysed the performance of different modalities depending on the distance of the camera to the person in order to assess the feasibility of a possible combined system using the various modalities supported by the proposed algorithm.

To validate the algorithm, a dataset has been recorded, composed of sequences natural colour, segmentation masks and depth sequences, in addition to recording of the heart rate measured by a heart rate monitor. Using this dataset, a complete set of results in the different situations under study has been obtained.

This algorithm is a reproduction based on the algorithm developed by Balkrishnan at MIT, although there may be modules that are not exactly equally implemented, because they were not described with all the details of how they were implemented originally.

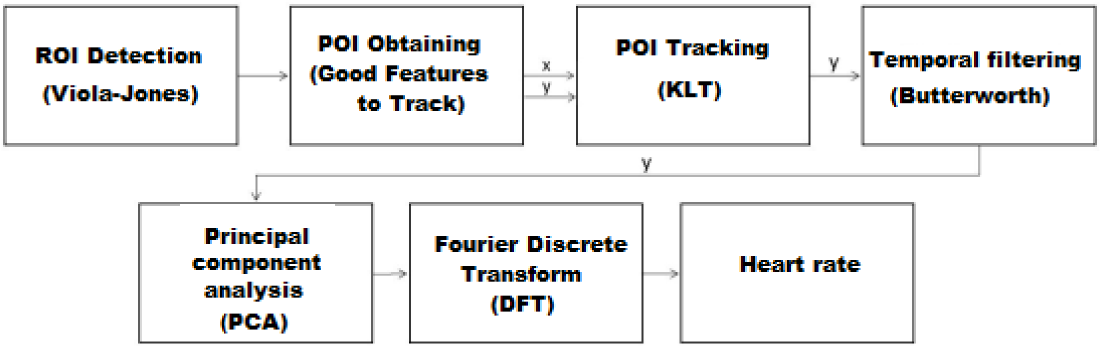

Figure 1: Base algorithm diagram.

As seen in the diagram, this algorithm is based on obtaining pulse induced by blood flow in the head movement. We have followed the scheme of Figure 1. First the region of interest is obtained using the Viola-Jones algorithm, later get the minutiae using Good Features to Track. The algorithm Good Features to Track, based on the Shi-Tomasi corners detection algorithm, works under certain assumptions that make more stable detected points for tracking. Points found are tracked with KLT. The signals obtained as a result of tracking is filtered with a Butterworth bandpass order 5 and passband between 0.75 Hz to 5 Hz, and they applied the principal component analysis (PCA) on the remaining components in the range between the second and fifth. To finalize the Discrete Fourier Transform (DFT) is performed on this energy to search the frequency at which the maximum is obtained and the result of this is considered the heart rate.

3. People privacy preservation in video-surveillance

This work consists on studying the implementation of a privacy preserving module for the existing video surveillance systems [2]. The unstoppable growth of these systems in many scenarios of our daily life implicate the development of techniques that permit hiding personal features of those individuals involved in tracking video sequences, providing that original sequence can be restored for forensic use after corresponding judicial authorization.

Firstly, an exhaustive study of state of the art is performed. Existing privacy preserving techniques, those reversible and those that are not, are detailed. Furthermore, a scenario in which this privacy module can be integrated in is also presented. After this, the technique that respects privacy, reversibility and coding losses' boundaries is selected.

Later, an algorithm that performs the selected technique is programmed. This algorithm is evaluated with several person detectors in order to study the behavior, introducing alternatives to improve its performance as well. Once best possible results are achieved, this algorithm is integrated in an application that detects persons after introducing the chosen privacy technique and shows visual real-time algorithm's features using some basic functions.

To change the sign of the DCT coefficients according to a PRNG (Pseudo Random Number Generator) is proposed in the state of the art. From an initial seed, a bit sequence of ones and zeros randomly determines which coefficients change sign and which is not generated. The initial value or seed is encrypted using a key and included in the stream to undo the scrambling in the decoder. In addition, since the scrambling is done in the transformed domain and before the formation of the encoded bitstream, any conventional decoder can undo the scrambling of the encoded bitstream without altering it. The DC coefficients are always positive, so you can not consider a random alteration sign change. However, for "changing the sign" according to a certain threshold, the half of the maximum value that could take the DC coefficient could be taken.

Although scrambling applying AC coefficients provides good levels of security it can also encrypt the DC coefficients altering the quantized value. In any case, the compression efficiency does not drop significantly and the computational complexity is not increased.

Finally, note that this algorithm the scrambling is just applied to the top of the person (facial face a greater or lesser extent). Since the evaluation tries to check that the person detection techniques still work once the scrambling has been performed, so the smaller region scrambling, the lower impact on the people detections.

(a)

(b)

Figure 2: Original image (a) and visual result of applying the scrambling function (b).

4. Material identification through the Kinect technology

This work has as main objectives to design and capture an image database and a system capable to train a model and classify a set of previously captured images [3]. The aim of this process is to develop a prototype that performs an approximation to automatic material recognition through the Kinect sensor.

In order to achieve these objectives several stages need to be first fulfilled. A study about recent versions of the Kinect sensor is first required. Then, existing methods for material characterization should be reviewed. These can be understood as versions of Bidirectional Reflectance Distribution Function (BRDF). In particular Histogram of Oriented Gradients (HoG) can be seen as a two-dimensional approximation to BRDF. The generation of knowledge models via Support Vector Machines (SVM) is also analysed. Finally datasets currently available for research have been also listed.

A new dataset has been recorded using Microsoft Kinect capturing material selection from different capture and illumination incidence angles. In order to include additional information to colour images, depth and infrared images would be also included in the dataset. Each material image has an associated patch which aim is to isolate the material from its background.

Once the dataset has been recorded a system using GDF-HOG and EigHess-HOG descriptors in order to extract patch characterization has been design. These descriptors are rotation invariant HoG extensions. Through SVM knowledge models have been generated. These models allow to predict and classify untrained input instances.

Finally, system performance has been evaluated achieving good results in supervised situations and promising results in real scenes.

The proposed system has been implemented by designing a diagram showing the flow system by which we perform the automatic identification of materials and which we can visualize in Figure 2.

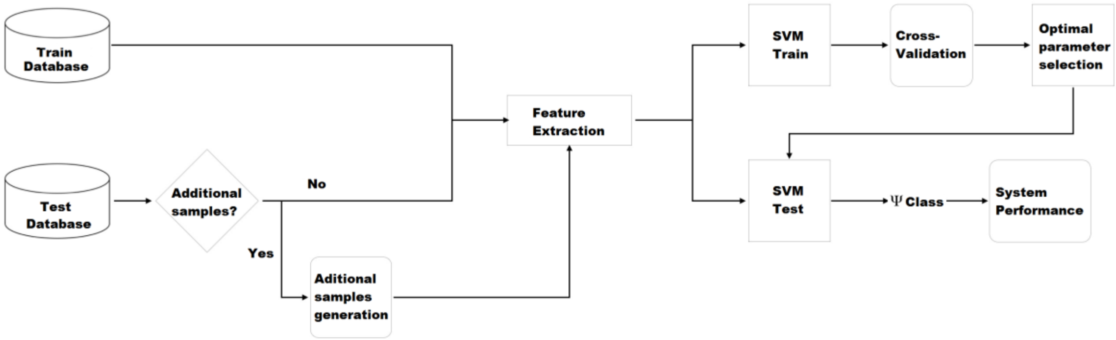

Figure 2: Flowchart of the proposed system.

This system will consist of two main branches: A training branch (Top of the diagram) and a test branch (bottom). Both paths are designed in the same manner but except the feature extraction block will be different.

Training Branch: in this branch of the flow chart we will train a knowledge model which will serve to predict future samples. This process of base training data instances to which we extract, by using descriptor, a feature vector, which will serve to train a Support Vector Machine (SVM).

Test Branch: starts from a database of test instances. These instances are used to the final measure system performance. The feature extraction block will be identical to the training branch (except for a replication module data). Then it makes use of model knowledge generated in the training branch and allow to predict test instances by SVM. These machines return a class (cotton, glass, …) for each input feature vector that we can use to measure performance of our system.

5. Design of support tools for the detection of logos in video sequences

This work describes an automatic method for the detection of brand logos in sport sequences. The work starts by studying the solutions existing in the state-of-the art in the topic. From this study a set of conclusions is derived, and these are used to define the design of the proposed method.

The method starts by defining three pre-processing methods which—motivated by design-heuristics—determine the spatial areas on which a logo is prone to be placed. Specifically, the methods use colour, structural and saliency based strategies to constrain the areas on which the logo detection process takes place.

On the candidate areas—those prone to contain a logo—, a classical point-of-interest matching strategy is used to relate the candidate instances with a preload logo template. From these matches, an affine correction of the template is derived. Logos are detected by measuring the similarity between the transformed template and the candidate areas.

Experimental results on a set of candidate sequences partially validate the design and development of the method for soccer sequences. However, results also illustrate the method's drawbacks and limitations when analyzing sequences of alternative sports. Furthermore, preliminary experiments on the use of the method for the generation of publicity statistics are also included, obtaining promising results. In overall, results suggest that the use of pre-processing techniques may help in the task of automatic logo detection.

Figure 3: Example of a sequence frame in which the detected logos are shown (green boxes).

6. Material identification through the Kinect technology (continuation)

The objective of this project is to use frequency information in a sequence of images to be able to automatically identify materials [5]. In addition, it also seeks to study the impact on this task including images with infrared light and depth information (enhancing RGB information).

To do this, the authors start by creating a new dataset with the help of the Kinect sensor, composed of RGB, depth and infrared images. The recorded sequences are divided into two parts: one part for training and another part for testing. Those intended for the training of the model are manually labelled to separate the materials from each other.

Once the materials are separated, descriptors are generated using this information and the Fourier transform, and knowledge models from these data are created, using the probabilistic model Gaussian Mixture Models, obtaining a multidimensional model for each material.

In the evaluation part the goodness of the models is studied using the sequences assigned to the evaluation part. The Fourier transform of the complete sequence is performed and introduced in each of the generated models, so that it returns the probability that each pixel has to belong to an object. A study on potential pre-processing methods on the analysed data has been also carried out. Results of the developed characterization methods are promising for the limited, yet generalist, subset of materials studied.

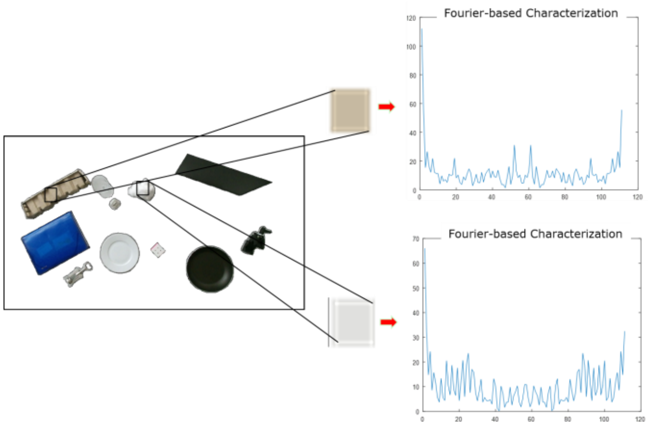

Figure 4: Fourier based characterization process for material identification: The temporal vibrations on a given square patch are stored for a given information (colour/ depth / infra-red). The characterization for each patch is obtained by the Fourier transform of such temporal variations.

7. Application of multi-plane augmented reality using correspondences between points of interest

The objective of this project is to create a naïve prototype of augmented reality [6]. To this aim, the tools and concepts of computer vision, machine Learning and the use of graphics are combined. Together, the shape an application that can detect an object and its pose and generate augmented reality on it. After studying the different possibilities for the creation of applications of augmented reality, a technique without markers (marker-less) have been selected.

In order to fulfil the objectives, the design of the prototype has been arranged on a per stage basis:

1. The detection of points of interest for the detection and description of objects.
2. The training of a support vector machine.
3. The use of computer vision techniques.
4. The use of a graphic engine.

The model has been evaluated and the results are, in general, promising. However, limitations have been found when detecting the object subjected to heavy rotations. The results of the developed prototype are positive and motivate further research.

Figure 5: Qualitative results of the Augmented Reality prototype. A simple 3D object is overlaid on a real object. Performance decreases with capture scale of the real object.

8. Conclusions

This document has described different research lines. The six works presented show promising results, using different techniques in each of the systems. All these technologies may contribute additionally to the project technologies.

Future research alternatives not directly considered within the project that may contribute additionally to the project technologies will be considered. It may comprise the use of additional sensors, processing methodologies, or the study of new application areas where the developed approaches may be of use (technological re-usability) for other project main research lines.

References

[1] TFG: Detección de ritmo cardíaco mediante vídeo (Heart rate detection using video), Erik Velasco Salido (advisor: José M. Martínez), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jun. 2015.

[2] TFG: Preservación de privacidad de personas en vídeo-seguridad (People privacy preservation in video-surveillance), Jaime Mateo Herrero (advisor: José M. Martínez), Proyecto Fin de Carrera (Master Thesis), Ingeniería de Telecomunicación, Universidad Autónoma de Madrid, Escuela Politécnica Superior, Oct. 2015.

[3] TFG: Identificación automática de materiales usando el sensor Kinect (Material identification through the Kinect technology), Alejandro López Cifuentes (tutor: Marcos Escudero Viñolo), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jun. 2015.

[4] TFG: Diseño de herramientas de apoyo para la detección de logotipos en secuencias de video, Efrén Martín García (tutor: Marcos Escudero Viñolo), Trabajo Fin de Grado, Grado en Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jul. 2016.

[5] TFG: Identificación de materiales utilizando el sensor Kinect V2(Material identification through the Kinect technology V2). Miguel Basarte Mena (tutor: Marcos Escudero-Viñolo), Trabajo Fin de Grado (Graduate Thesis), Escuela Politécnica Superior, Universidad.Autónoma de Madrid, Jul. 2017.

[6] TFG: Aplicación de realidad aumentada multi-plano mediante correspondencias entre puntos de interés (Application of multi-plane augmented reality using correspondences between points of interest). Alejandro Nuñez Valle (tutor: Marcos Escudero Viñolo), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jul. 2017.