





ITN-DCH – Initial Training Network for Digital Cultural Heritage

Projecting our Past to the Future

FP7-PEOPLE ITN2013 - Project Contract Number 608013

Deliverable number	3.4
Title	Object Tracking in Digital Cultural Heritage

Due Date	September 2016
Actual date of delivery to EC	

Included (indicate as appropriate)	Executive Summary	☒	Abstract	☒	Table of Contents	☒

Project Coordinator:
Dr. Marinos Ioannides
Cyprus University of Technology (CUT)
Digital Heritage Research Lab
Tel. +357-25-002020
E-mail: marinos.ioannides@cut.ac.cy
Project website address: www.itn-dch.eu



Context:

Partner responsible for deliverable	KU Leuven & Fraunhofer
Deliverable author(s)	Gina Stavropoulou, Luc van Gool
Deliverable version number	1

Dissemination Level	
Public	☐
Restricted to other programme participants (including the Commission Services)	☐
Restricted to a group specified by the consortium (including the Commission Services)	☐
Confidential, only for members of the consortium (including the Commission Services)	☒

History:

Change log			
Version	**Date**	**Author**	**Reason for Changes**
0.1	05.07.2016	Gina Stavropoulou	Introductory chapters and template
0.2	01.09.2016	Gina Stavropoulou	First Draft
0.3	03.09.2016	Gina Stavropoulou	Edits

Release Approval			
Version	**Date**	**Name-Organisation**	**Role**



Statement of Originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgment of previously published material and of the work of others has been made through appropriate citation, quotation or both.



CONTENTS

1 Executive Summary **6**

2 Introduction **7**
2.1 Role of this Report in the Project . . . 7
2.2 Approach and Objectives . . . 7
2.3 Structure of the Document . . . 7

3 Project Deliverables, Milestones, and Workpackages **8**
3.1 Project Deliverables . . . 8
3.2 Milestones . . . 9
3.3 Work Packages . . . 10

4 Objectives of Work Package 3 and the Deliverable 3.4 **11**

5 Achievements, Risks and Issues **12**

6 Fellows Involvement **13**

7 Full Partners Involvement **14**

8 Associate Partners Involvement **15**

9 What is Object Tracking? **16**
9.1 Object Representation . . . 20
9.2 Feature Extraction . . . 22

10 Applications in the Cultural Heritage Domain **24**
10.1 Motion Capture . . . 24
10.2 Eye tracking . . . 25
10.3 Vuforia 3D object tracker . . . 26

11 Impact of Any Deviation from the Annex I **28**

12 Next Steps **29**

References **30**



LIST OF FIGURES

4.1 Diagram showing the connections between the ITN-DCH work packages 11
9.1 Face detection [Source: (Rowley, Baluja, & Kanade, 1998)] 16
9.2 Pedestrian detection [Source: (Dollár, Wojek, Schiele, & Perona, 2009)] 17
9.3 Vehicle detection and trajectory tracking. [Source: (Teutsch & Grinberg, 2016)] 17
9.4 Face detection based on skin colour detection. [Source:(Hasan, Thakur, & Podder, n.d.)] 18
9.5 Taxonomy of Tracking Methods. [Source:(Yilmaz, Javed, & Shah, 2006)] 18
9.6 Object representations. (a) Centroid, (b) multiple points, (c) rectangular patch, (d) elliptical patch, (e) part-based multiple patches, (f) object skeleton, (g) complete object contour, (h) control points on object contour, (i) object silhouette. [Source:Yilmaz et al. (2006)] 20
9.7 Skeletonization of a Japanese symbol and a handwritten word for optical character recognition applications. [Source:Saeed, Tabędzki, Rybnik, and Adamski (2010)] 21
9.8 Example of human pose estimation based on articulated models. [Source:Yang and Ramanan (2011)] 21
9.9 Examples of different image features used for object tracking as shown in Luo, Xing, Zhang, Zhao, and Kim (2014). (a) Image showing optical flow (S. Ali & Shah, 2008), (b) image showing covariance matrix, (c) image showing point feature (Brostow & Cipolla, 2006), (d) image showing gradient based features (Dalal & Triggs, 2005), (e) image showing depth (Mitzel, Horbert, Ess, & Leibe, 2010) and (f) image showing color feature (Okuma, Taleghani, De Freitas, Little, & Lowe, 2004). 22
10.1 Motion capture session with a dancer [credit: MIRALab] 24
10.2 Examples of different perception of the same image depending on the task required. [Source:Yarbus (1967)] 25
10.3 Examples of an object being scanned using a printed target. 26
10.4 Example of Vuforia being used for the Carnuntum case study. 27



1 EXECUTIVE SUMMARY

Initial Training Network for Digital Cultural Heritage: Projecting our Past to the Future (ITN-DCH) is a Marie Curie fellowship research programme under the FP7 PEOPLE framework. The current document represents the Deliverable 3.4, which is part of the Working Package 3 with title "Object Tracking in Cultural Heritage" and with KUL acting as a leader and Fraunhofer acting as co-leader.



2 INTRODUCTION

2.1 ROLE OF THIS REPORT IN THE PROJECT

ITN-DCH and the wider field of Digital Cultural Heritage are characterised by multi-disciplinarity, as several researchers from different fields and disciplines have to co-operate and combine their efforts on a common final goal leading to scientifically complete and accurate outputs. However, the main problem in this field, as in any other multi-disciplinary field, is the lack of a common language between the scientists. It is a frequent phenomenon that archaeologists or museum managers do not understand the potential hidden in engineering/computer science or that their requested tasks that are not achievable. On the other hand, engineers/computer scientists have limited knowledge on what are the expectations of digital cultural heritage applications. Recognising this problem, ITN-DCH aims to bridge this communication gap between archaeologists, architects, engineers, computer scientists and other research working within its frame. Therefore the current deliverable but also the rest of the deliverables of the WP3 aim to serve as a framework for the fusion of the two disciplines; Computer vision and Cultural Heritage.

2.2 APPROACH AND OBJECTIVES

Main objective of the WP3 is to develop geometrically and semantically enriched pre/post data processing, filtering and feature extraction methodologies for application to cultural heritage. Therefore the research for the specific WP is focused on the fields of computer vision, machine learning and data processing. More specifically the purpose of WP3 is to suggest methodologies of automating 2D and 3D content digitisation processing, through OCR and 2D/3D geometrically enriched features accompanied with visual matching and learning methodologies. To achieve this, widely used computer vision algorithms need to be re-examined under the light of cultural heritage applications.

2.3 STRUCTURE OF THE DOCUMENT

The document is structured based on the ITN-DCH template for reports.



3 PROJECT DELIVERABLES, MILESTONES, AND WORKPACKAGES

ITN-DCH project deliverables, milestones and work packages as described in the Annex I are listed below (Tables 3.1, 3.2, 3.3).

3.1 PROJECT DELIVERABLES

Del. Nr.	Deliverable Title	Lead Beneficiary (acronym)	Delivery Due Date	Submission Date
0.1	Management Reports		3	
1.1	Initial Cultural Heritage Requirements and scenario description from end-users' point of view		12, 24, 36, 48	
1.2	Survey of current specifications used in CH repositories		12	
1.3	Updated Cultural Heritage Requirements and scenario description from end-users' point of view		14	
1.4	New metadata interface for tangible cultural heritage (3D/4D forms)		18	
1.5	New metadata interface for intangible cultural heritage (3D/4D forms)		30	
1.6	The final Digital Cultural Heritage Description Interface		36	
2.1	Design of the digitalization architecture		40	
2.2	Camera calibration and Automatic registration of large data set		12	
2.3	Data acquisition techniques using multi-view cameras		18	
2.4	Data acquisition techniques using laser scanners and depth sensors		24	
2.5	Visual fusion methodologies and super-resolution methods		27	
3.1a, 3.1b	Geometry enriched feature extraction algorithms		20, 28	
3.2	Image processing and analysis for large data set filtering		28	
3.3	3D modelling, and reconstruction		33	
3.4	Object tracking in Digital Cultural Heritage		36	
4.1	Learning and pattern recognition algorithms		24	
4.2	Semantic representation and ontology-based conceptualization		36	
4.3	Symbolic representation		40	
5.1	First version of architectural design		24	
5.2	Final architectural design		30	
5.3	4D integrated cultural heritage viewer (GIS, time and animation)		40	
5.4	Mixed and augmented reality environments		40	
5.5	Media production, and social services		42	
5.6	Real-time and perceptual rendering		44	



6.1	QoS and cloud computing architectures for delivering real- and interactive multi-media services		48	
7.1	Dissemination Outreach plan		8	
7.2	Exploitation Plan		14	
7.3	Workshops		6. 15, 26,32,40	
7.4	Final outreach and dissemination actions		48	
7.5	Standardization		48	

3.2 MILESTONES

WPs	Milestones	Milestone	Lead Beneficiary	Month	Completion Date
WP0	ML1	Recruitments strategy and recruitments phase completed	CUT	4	
WP8	ML2	Outreach, dissemination and exploitation plan completed and validated	CUT	8	
WP1-5	ML3	First set of requirements completed on research-through-training work packages WP1-5	UMU	18	
WP1-5	ML4	First phase of research-through-training progress on WP1-5 completed	KUL	24	
WP1-5	ML5	Requirements and specifications completed, updated and validated for all research-through-training work packages	CAD	24	
WP1-5	ML6	Second phase of research-through-training progress completed on WP1- 5	IGD	34	
WP5	ML7	The next generation digital cultural heritage services specified	MIRALab	34	
WP1	ML8	The final Digital Cultural Heritage Description Interface (for tangible and intangible content) validated by the end-users and private sector	UMU	42	
WP5	ML9	Third phase of research-through training progress	FORTH	42	
WP7	ML10	Exploitation activities completed	7Reasons	46	
WP7	ML11	Outreach and dissemination activities completed	NTUA	48	
WP5	ML12	Validation of digital cultural heritage services completed	ArcTron	48	



3.3 WORK PACKAGES

WP	Work Package Title	Type of Activity	Lead (Co-Leader/Backup) Participant	Other Participants Involved	Start Month	End Month
WP0	Management	Management	CUT (IGD)	ALL	M1	M48
WP1	Cultural Heritage Data Definitions, Requirements and the Metadata Description Interface	Research/ training	UMU (CAD)	NTUA, UL, UNIZAG, RMN, MC, KAAK	M5	M40
WP2	Capturing and Digitization technologies in Cultural Heritage	Research/ training	ArcTron (NTUA)	USt , FBK, CAD, UMU, RMN, MC, KAAK	M5	M30
WP3	Data pre/post-processing (2D, 3D), Computer vision and computational geometry	Research/ training	KUL (IGD)	NTUA, USt, ArcTron, FBK, CAD, UMU, RMN, MC, KAAK	M9	M36
WP4	Semantic-Symbolic & Conceptual representation	Research/ training	FORTH-semant (CNRS)	IGD, USt, UL, NTUA, UNIZAG, eSelis	M13	M40
WP5	Digital Cultural applications	Research/ training	MIRALab (FORTH-augm)	IGD, UW, ArcTron, 7Reasons, eSelis	M18	M48
WP6	Cloud Services in Digital Cultural Heritage:	Research/ training	USt (NTUA)	IGD, ArcTron	M25	M44
WP7	Dissemination, Outreach, Exploitation and Standardization Actions	Dissemination, outreach	UW (FBK)	ALL	M1	M48

4 OBJECTIVES OF WORK PACKAGE 3 AND THE DELIVERABLE 3.4

Work Package 3 with title "Data pre/post-processing (2D, 3D), Computer vision and computational geometry" started on the 9th month of the project and is expected to finish on the 36th. KU Leuven acts as a leader for the specific WP, organising and distributing tasks to the involved partners, whereas FRAUNHOFER has the role of the co-leader, offering support in the coordination.

In the diagram below [Fig.4.1] it is obvious that data pre/post processing and computer vision are not stand-alone fields but they are dependent on the data acquisition techniques, they require clear definition of metadata and they form a challenge for the field of semantic representation. To this end, WP2 and WP3 are bi-directionally related with each other and additionally WP4 and WP1 are explicitly related to WP3.

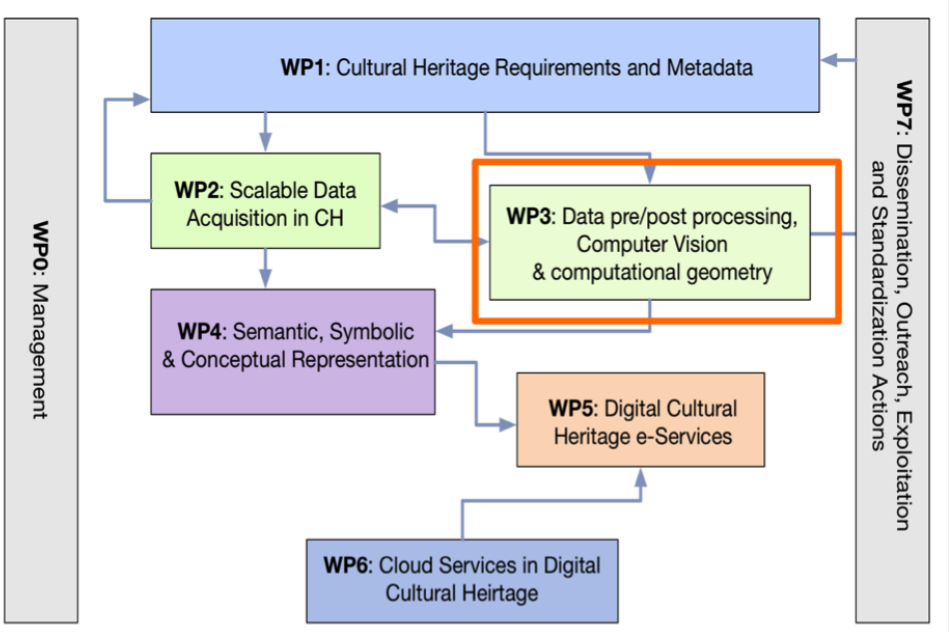


Figure 4.1: Diagram showing the connections between the ITN-DCH work packages

With regards to the novel approaches of this work package, the deliverables will explore the possibility of automating the modelling process for both tangible and intangible content and in cases of complex background, moving objects and severe occlusions. However, most of the work regarding these deliverables is still in progress.



5 ACHIEVEMENTS, RISKS AND ISSUES

There are no risks relevant to the current deliverable.



6 FELLOWS INVOLVEMENT

KU Leuven, partner of the ITH-DCH project, is leader of the Working Package 3 ("Data Pre/post-Processing - 2D, 3D Computer Vision and Computational Geometry") and Fraunhofer acts as a co-leader. The following table shows the involvement of the fellows of the project in the deliverable 3.4, based on the Annex I.

Table 6.1: Fellows Involvement (credit: ITN-DCH).

ESR	Partner	Must Participate to Deliverable 3.4	Relevant for the underlining task 3.4
ESR1	USTUTT		
ESR2	FBK		✓
ESR3	ArcTron		✓
ESR4	NTUA-PhotoLab		
ESR5	KUL	✓	✓
ESR6	UL		
ESR7	IGD	✓	
ESR8	CNRS		
ESR9	7Reasons		✓
ESR10	FORTH		
ESR11	UMU		
ESR12	CUT		
ESR13	MiraLab		
ESR14	UL		
ESR15	NTUA- Materials		
ESR16	UW		
ER1	ICS-FORTH		
ER2	CUT		✓
ER3	CUT		
ER4	MiraLab		



7 FULL PARTNERS INVOLVEMENT

Several ITN-DCH partners are involved in current deliverable directly or indirectly. The table below (Table 7.1) gives an overview of all partners should contribute in general in the WP3 and subsequently in the deliverable 3.4.

Table 7.1: Full Partners Involvement in WP3 (credit: ITN-DCH)

	Full Partner	Relevant Tasks
ACADEMIA	KUL	Computer vision for extracting geometrically enriched features.
	NTUA	Image processing algorithms for noise removal; 3D data filtering; machine learning for filling possible gaps and transforming point clouds of different orientations to a common coordination system.
	UW	Pre (post)-processing in relation to the day-light reflections & environments
	MiraLab	3D/4D reconstruction, simulation and texture mapping.
	USTUUT	GIS Images and pre-processing of complex data sets.
RESEARCH	FRAUNHOFER	Direct pre-processing of different data attributes such as colours on frescos, paintings, 3D/4D data.
	FORTH	Computer graphics for real-time gesture tracking and recognition.
INDUSTRY	ArcTron	Standardized software platforms for data processing, filtering, filling of the gaps, texture rendering on cultural heritage sites data sets.



8 ASSOCIATE PARTNERS INVOLVEMENT

The following table (Table 8.1) indicates the Associate partners that should be involved in the WP3 and subsequently in the current deliverable.

Table 8.1: Associate Partners Involvement in WP3.

Associate Partner	Relevant Tasks
KAAK	Training on 3D/4D GIS & Modelling
RMN	Tailored-made data pre-processing
eSelis	OCR, text and image processing/formatting

9 WHAT IS OBJECT TRACKING?

Object tracking is an important task for many computer vision applications that require video understanding and its main purpose is to detect and keep track of the trajectory of a moving object through a sequence of images or a video along with other kinds of information such as shape, area and orientation. The terms object tracking and object detection are often confused or used interchangeably, however, object detection techniques are usually used as a first step in object tracking pipelines and it is a task that is applied on single images or on the individual frames of a video. The most common object detection tasks involve the detection of faces (Li, Lin, Shen, Brandt, & Hua, 2015; Rowley et al., 1998) [Fig. 9.1], humans/pedestrians (Dollár et al., 2009) [Fig. 9.2] or vehicles. The detection of the desired objects in one image and their association with the corresponding objects in other images are two tasks that can be performed either separately or jointly (Yilmaz et al., 2006). In the first case an object detection algorithm is applied in each frame and then a tracker is used to associate it with the corresponding object in the sequence of frames. Alternatively, in the second case, both the object and its correspondence in other frames are iteratively estimated and simultaneously updated based on information from previous frames (Yilmaz et al., 2006).



Figure 9.1: Face detection [Source: (Rowley et al., 1998)]



Figure 9.2: Pedestrian detection [Source: (Dollár et al., 2009)]

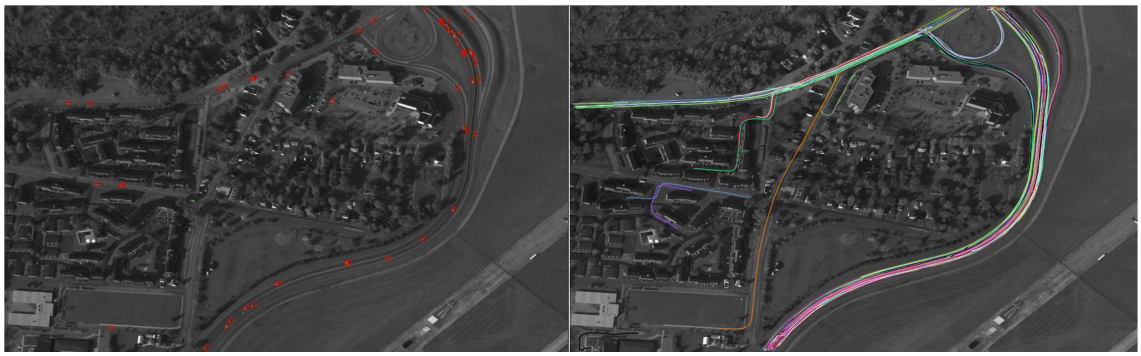


Figure 9.3: Vehicle detection and trajectory tracking. [Source: (Teutsch & Grinberg, 2016)]

Although researchers have worked intensively on object detection and tracking for many decades, these research fields still remain active and challenging. As in most computer vision problems, there is not an one-method-fits-all solution but rather a wide range of different solutions targeted to the different challenges that can arise. The difficulties of these tasks can be attributed to non-continuous or complex object motion,deformations, non-rigid object structures, image noise or illumination changes, unstable or cluttered background, object-to-object and object-to-scene occlusions and camera motion. Additionally, since the object tracking can be a computationally intensive task there is a need for real-time processing methods. Since the challenges for object tracking applications are numerous, usually the current approaches impose certain constraints on the object; its motion should not be abrupt or that the velocity should be stable. Additional constraints could involve predetermined size or number of object to be detected. Therefore, the representation of the object to be detected, i.e. the group of features which are used to describe the object, can significantly affect the tracking pipeline. If only a specific colour is used to describe the objects then the detection and subsequently the tracking can be significantly simplified. An examples of colour based detection would be face and person detection based on skin detection (Hasan et al., n.d.) [Fig.9.4]. However, when the task is more complicated as in the case of human detection and the objects are characterized by more complex, higher level features adding to the problems of different poses and appearances, then their detection can be a very challenging task. More on object representations can be found in the following chapter. Due to the numerous aforementioned challenges, the complexity of the environment, the variety of possible scenarios and the number of different object representations and features, there are many different approaches to object tracking [Fig. 9.5]. Each different approach has been suggested to tackle a different issue and researchers today should choose the most suitable method according to the needs of their application.

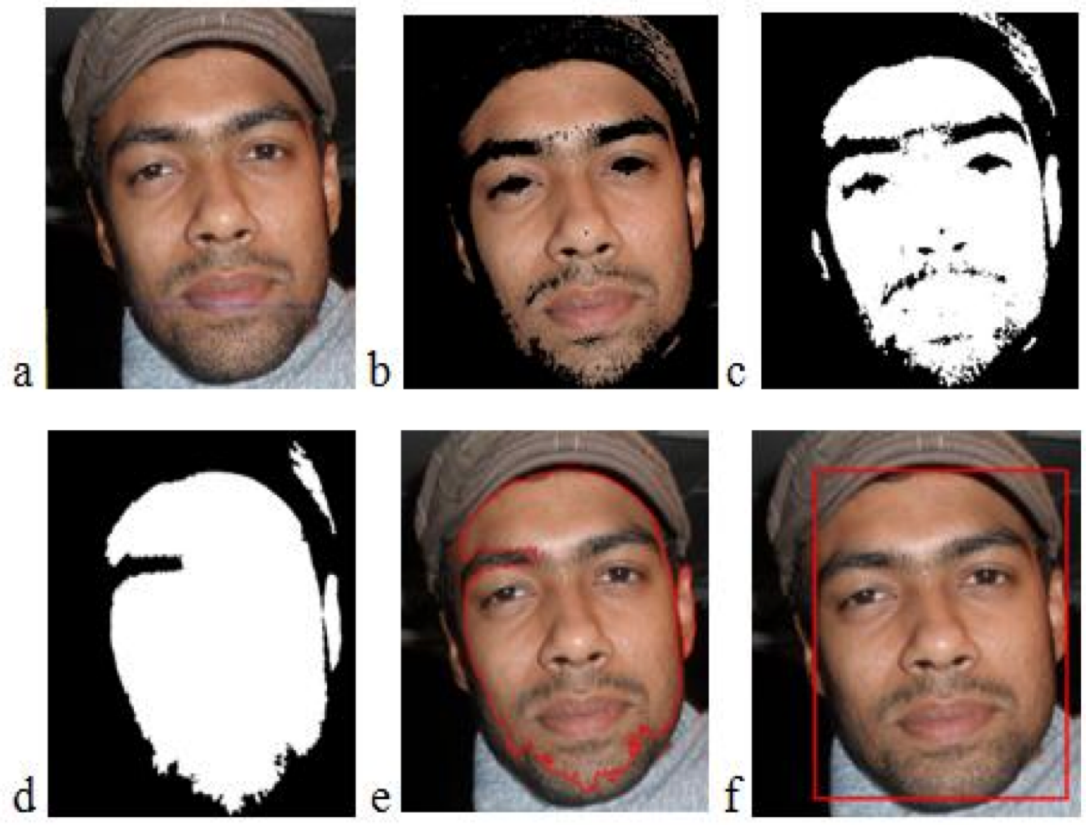


Figure 9.4: Face detection based on skin colour detection. [Source:(Hasan et al., n.d.)]

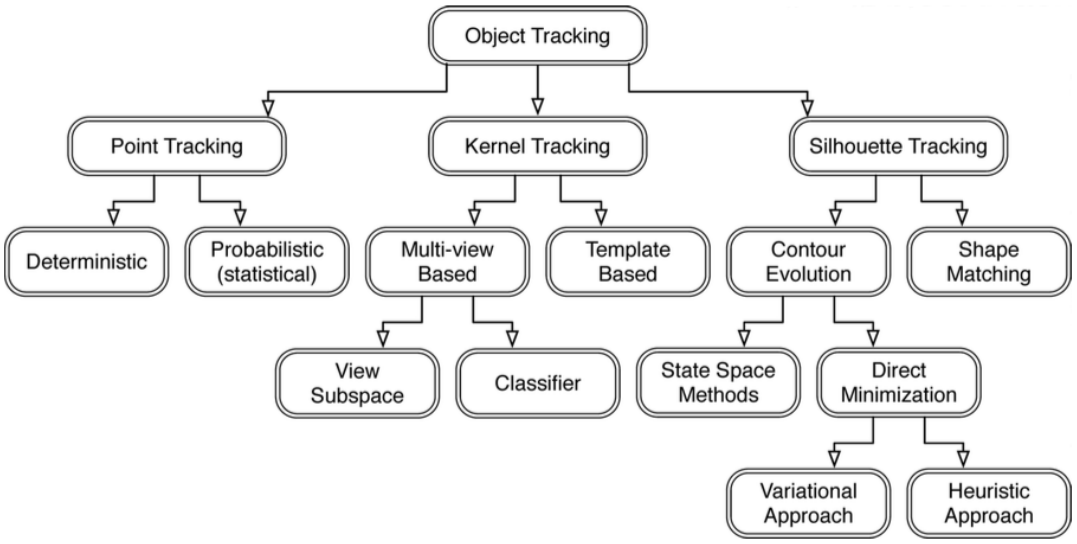


Figure 9.5: Taxonomy of Tracking Methods. [Source:(Yilmaz et al., 2006)]

Object detecting and tracking has a wide variety of applications in computer vision such as video surveillance (Xu, Zhang, & Goto, 2011; Kim & Davis, 2011), crowd (Dehghan & Shah, 2016; Dehghan &



Kalayeh, n.d.; Chan, Liang, & Vasconcelos, 2008) or traffic monitoring (Kiratiratanapruk & Siddhichai, 2006), human-computer interaction applications (Prisacariu & Reid, 2012; Jacob & Karn, 2003), medical imaging (Cao, Li, & Sun, 2015; Wang et al., 2013) and augmented reality applications (Uchiyama & Marchand, 2012; Gammeter, Gassmann, Bossard, Quack, & Van Gool, 2010; Park, Lepetit, & Woo, 2008). Additionally, it provides input to higher level vision tasks, such as 3D reconstruction and 3D representation, usually in the form of feature tracking. Feature tracking is a fundamental problem in computer vision with applications in various tasks including 3D reconstruction and visual SLAM (Ji, Li, Salzmann, & Zhong, 2016; Shi & Tomasi, 1994).

In the field of cultural heritage there have been many interesting applications of object tracking. Apart from using object tracking systems for surveillance and protection from theft, the most interesting applications involve motion capture for the documentation of intangible heritage, eye tracking methods for studying art appreciation and of course, object tracking for augmented reality applications. All these applications will be discussed in the following chapters.

9.1 OBJECT REPRESENTATION

Object tracking can have many different applications in different domains. Therefore the objects to be detected can differ significantly, from humans to different kind of animals or vehicles to blood cells, depending on the tracking scenario. However,as in most computer vision tasks, the objects under study are usually represented by employing certain 2D shape representations. Many different representation models have been proposed [Fig.9.6], each serving different purposes and tackling different challenges. Below we see some of these 2D representations, as described by Yilmaz et al. (2006).

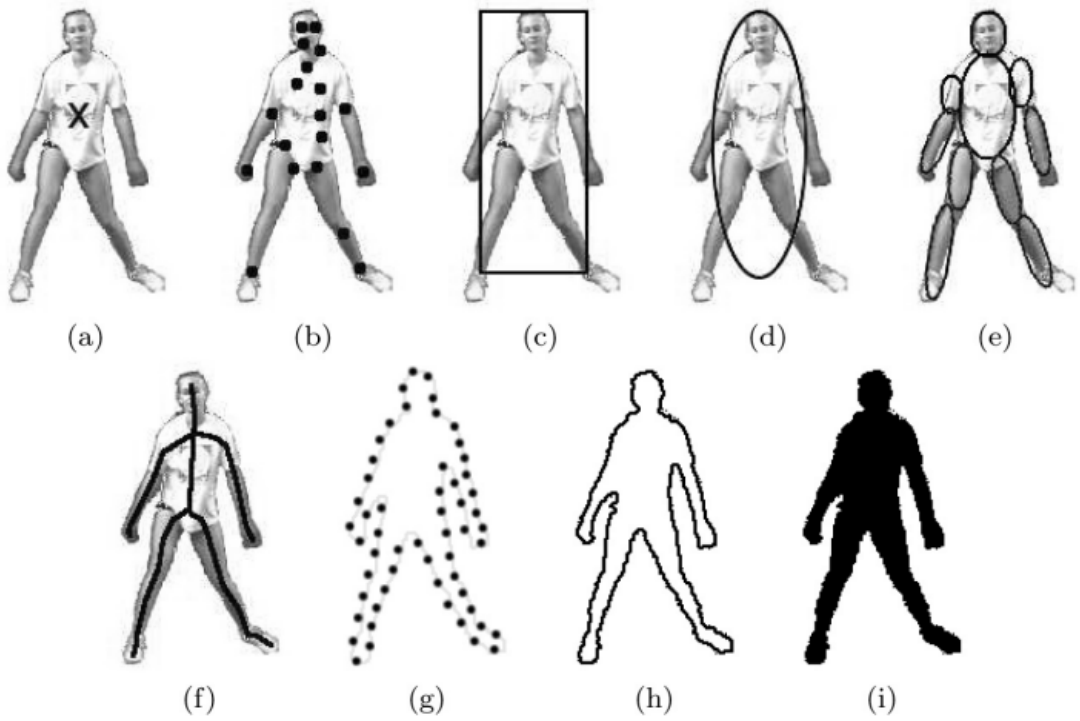


Figure 9.6: Object representations. (a) Centroid, (b) multiple points, (c) rectangular patch, (d) elliptical patch, (e) part-based multiple patches, (f) object skeleton, (g) complete object contour, (h) control points on object contour, (i) object silhouette. [Source:Yilmaz et al. (2006)]

- *Point Representation:* An object can be represented by a single point, usually its centroid [Fig. 9.5(a)] /citeveenman2001resolving or by a set of points [Fig. 9.5(b)] (Serby, Meier, & Van Gool, 2004). This representation is usually used for objects that have a small area in comparison to the image area, like small blobs.
- *Primitive geometric shapes:* The shape of the object can be represented by one of the primitive geometric shapes, i.e. rectangles, circles, triangles, ellipses etc [Fig. 9.5(c),(d)] (Comaniciu, Ramesh, & Meer, 2003) or a group of such shapes. The motion of the object can then be modeled by translation, affine, or 2D projective transformation. Though primitive geometric shapes are more suitable for representing simple rigid objects, they are also used for tracking non-rigid objects.
- *Object silhouette and contour:* The contour representation defines the boundary of an object with either a discrete number of nodes or with a continuous line [Fig. 9.5(g), (h)]. The region inside the

contour is called the silhouette of the object [Fig. 9.5(i)]. Such representations are more suitable for describing complex, non-rigid shapes (Yilmaz, Li, & Shah, 2004).

- *Articulated shape models:* Articulated objects are composed of parts that are held together with joints. For example, the human body is an articulated object with torso, legs, hands, head, and feet connected by joints and current research on human pose estimation frequently uses such models [Fig. 9.8] (Yang & Ramanan, 2011). The parts of an articulated object, are usually represented by cylinders,rectangles or ellipses as shown in figure 9.5(e).

- *Skeletal models:* Object skeleton [Fig. 9.5(f)] can be extracted by applying medial axis transform to the object silhouette [Fig. 9.7]. This kind of representation is frequently used for object detection, segmentation and recognition tasks (A. Ali & Aggarwal, 2001), as it provides a structural representation to describe the relationship among object parts and current research is still trying to find effective ways for object skeleton extraction in natural images (Shen et al., 2016).



Figure 9.7: Skeletonization of a Japanese symbol and a handwritten word for optical character recognition applications. [Source:Saeed et al. (2010)]

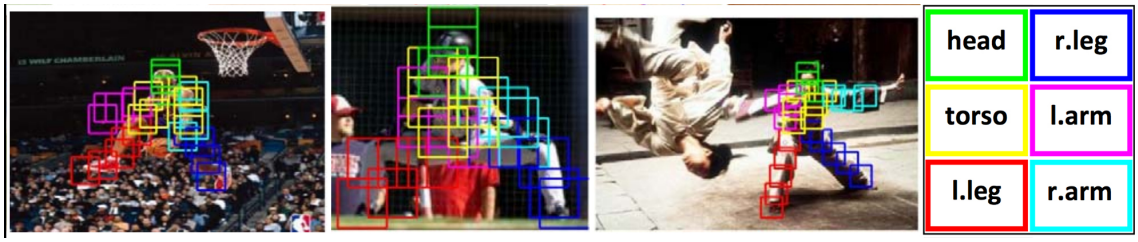


Figure 9.8: Example of human pose estimation based on articulated models. [Source:Yang and Ramanan (2011)]

9.2 FEATURE EXTRACTION

Apart from selecting the appropriate representation model for the objects to be detected, another important aspect of tracking is the image features. A good feature is the one that makes the object easily distinguished in the feature space. The appropriate features for each application can be selected based on the object representation. For example, if an object is represented by its contour, then the appropriate feature would be the edges and edge extraction algorithms should be applied on the image. In most tracking algorithms a combination of features is selected. Below a small categorization of image features is offered, as suggested by Luo et al. (2014) and Yilmaz et al. (2006). More on feature extraction can be found in the deliverable 3.1.a.

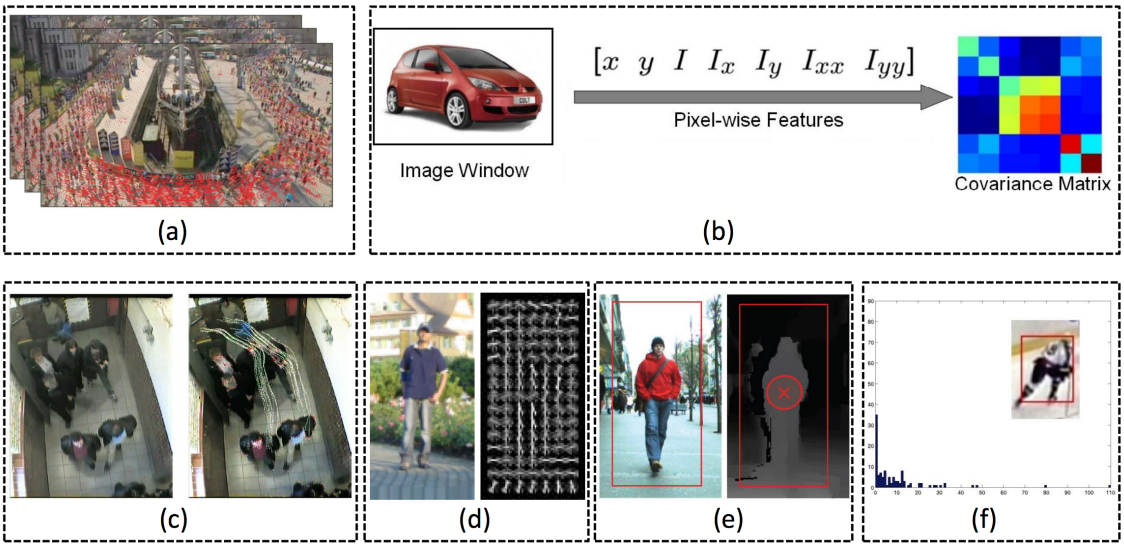


Figure 9.9: Examples of different image features used for object tracking as shown in Luo et al. (2014). (a) Image showing optical flow (S. Ali & Shah, 2008), (b) image showing covariance matrix, (c) image showing point feature (Brostow & Cipolla, 2006), (d) image showing gradient based features (Dalal & Triggs, 2005), (e) image showing depth (Mitzel et al., 2010) and (f) image showing color feature (Okuma et al., 2004).

- *Point and Edges:* Point features are mostly used for tracking of single objects (Shi & Tomasi, 1994), however they have also been used for multiple object tracking or for feature tracking. For instance, KLT tracker, as suggested by Tomasi and Kanade (1991), is employed to track feature points and generate a set of trajectories between the corresponding features. In order to extract points features point detectors are used and the reader is referred to deliverables 3.1.a and 3.1.b to find more information about them. For detecting object boundaries and contours, edges can be used as a feature. Therefore edge detection algorithms, such as the Canny edge detector (Canny, 1986), are often employed in tracking applications. An evaluation of the edge detectors has been performed by Bowyer, Kranenburg, and Dougherty (1999).
- *Colour:* This feature is frequently used in tracking applications using not only the RGB but a variety of colour spaces. Colour Histograms are also being frequently used as features (Okuma et al., 2004; Song, Jeng, Staudt, & Roy-Chowdhury, 2010; Mitzel et al., 2010; Izadinia, Saleemi, Li, & Shah, 2012).



- *Optical flow:* Optical flow is a feature very commonly used in tracking applications. It uses a dense field of displacement vectors that indicate the translation of each pixel between two consequent frames. This feature is applied under the assumption that the brightness of the object and background pixels remains unchanged in consecutive frames (Horn & Schunck, 1981). Many different methods for computing the dense field of vectors have been suggested (Horn & Schunck, 1981; Lucas, Kanade, et al., 1981; Black & Anandan, 1996; Szeliski & Coughlan, 1997) and an evaluation of them has been performed by Barron, Fleet, and Beauchemin (1994). Additionally, in more recent literature, optical flow has been utilized to associate the detection responses from continuous frames (Rodriguez, Ali, & Kanade, 2009; Izadinia et al., 2012) and also for discovering motion patterns in extremely crowded environments (S. Ali & Shah, 2008; Rodriguez, Sivic, Laptev, & Audibert, 2011).

- *Gradient/pixel-comparison features:* There are some features which are based on gradient or pixel comparison. HOG is one the most famous gradient based features (Dalal & Triggs, 2005) features and it has been employed many times for detecting objects or for data association (Izadinia et al., 2012; Kuo, Huang, & Nevatia, 2010; Breitenstein, Reichlin, Leibe, Koller-Meier, & Van Gool, 2009).

- *Region covariance matrix features:* The region covariance matrix (Porikli, Tuzel, & Meer, 2006; Tuzel, Porikli, & Meer, 2006) has been used as an illumination and scale invariant feature in many tracking applications and even for data association (?, ?) and appearence learning (Hu et al., 2012; Kuo et al., 2010).

10 APPLICATIONS IN THE CULTURAL HERITAGE DOMAIN

10.1 MOTION CAPTURE

As mentioned by Guerra-Filho (2005), motion capture is the process of recording the real-life movement of a subject as sequences of Cartesian coordinates in the 3D space. More specifically, motion capture is a research field that tries to accurately record and reproduce complex movements in an accurate and realistic manner and it has many applications in many different domains such as sports, education and, of course, computer animation for cinema, television, and video games. In the cultural heritage domain, motion capture has been used for the digitization of intangible heritage and more frequently for the documentation of dances [Fig. 10.1].

There are four different ways to capture the movements, depending on what type of sensors are available: Optical, Inertial, Magnetic or Mechanical. Optical systems record the motion based on multiple image sensors and object tracking techniques. First a set of 2D images from different cameras positioned around the subject is acquired. Then each camera detects the points placed on the subject. Finally, the 2D images are processed using a projection matrix. The result is a geometrical reconstruction of the markers in 3D space.

The markers can be passive (reflective) or active (emitting a unique digital signature). Passive optical systems use markers coated with a reflective material to reflect infrared light generated by IR LEDs around each camera. In order to detect only the bright reflective markers for sampling, the threshold of the camera can be tuned to avoid the detection of fabric, skin and unwanted objects. In general, a system calibration is required. For each two dimensional image captured, the average position of each marker is estimated using image processing tools to isolate the markers. Calibration is required to determine the relative position of the cameras and eventually correct for lens distortion. This is achieved using a dedicated object with a specific pattern of markers on it, placed in the center of the sensing area. The marker swap is a recurrent problem which reduces by using hundreds of cameras. Usually many cameras are required for full coverage of the subjects or in case of multiple subjects in the same scene. More about motion capture can be found in deliverable 2.1.



Figure 10.1: Motion capture session with a dancer [credit: MIRALab]

10.2 EYE TRACKING

Eye tracking is a a motion capture method tailored to detect the point and the trajectory of gaze based on the movement of the eyes relatively to the head, i.e. to identify where the person under study is looking when given a visual stimulus. This is usually achieved with the use of eye trackers, devices capable of recognizing eye positions and motions. Similarly to motion capture systems, the state of the art eye trackers deploy special infra-red sensors to detect the corneal reflections and subsequently the movement of the pupil. Current research is focused on developing eye tracking algorithms that can function in real-time on hardware with limited processing power, such as mobile devices (Markuš, Frljak, Pandžić, Ahlberg, & Forchheimer, 2014).

Eye trackers are often used not only to study the human vision system but also in the field of psychology to study cognitive functions as memorization, visual attention, problem solving and decision making (Rayner, 2009). Eye tracking methods have also been frequently used in marketing for web and interface usability and advertising and as input for human-computer interaction applications. In the field of cultural heritage eye tracking had been used to study how individuals perceive, understand and appreciate art by recording their visual activity in front of paintings (Quiroga & Pedreira, 2011; Massaro et al., 2012; Naspetti et al., 2016). Recently Calandra, Di Mauro, D'Auria, and Cutugno (2016) introduce E.Y.E.C.U. (read "I see you"), an eye tracking system employed to enhance the experience of visiting a museum with augmented reality. The gaze of the visitors is tracked by the system and when it lingers on a painting illustrative content appears on the walls around it. Additionally, there is an emotional analysis module which monitors the dilation of the visitors pupil under different stimuli.

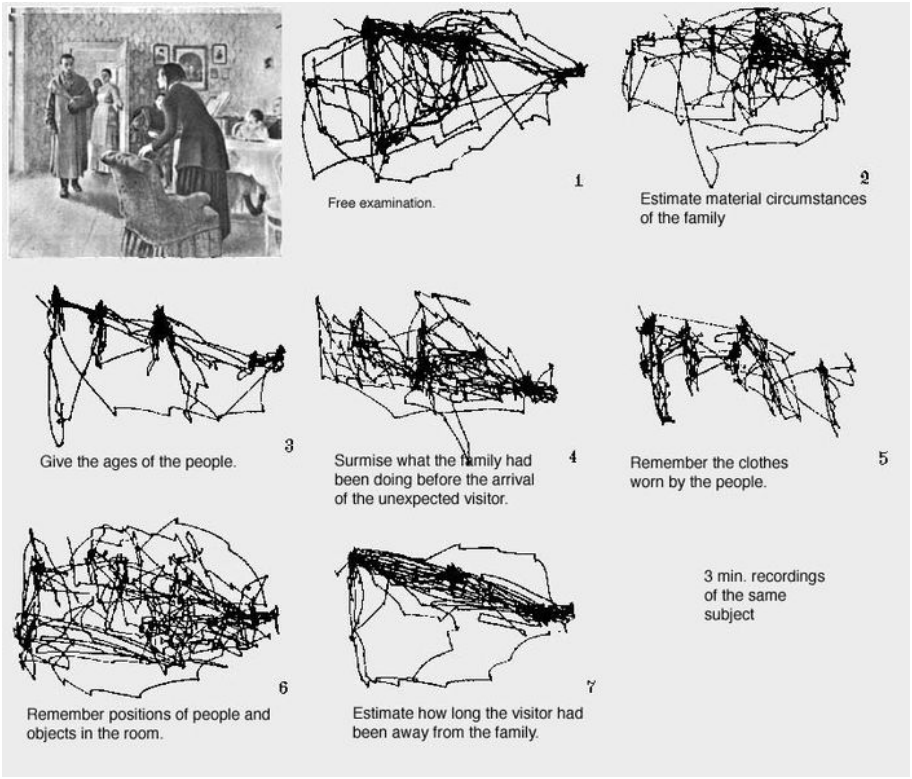


Figure 10.2: Examples of different perception of the same image depending on the task required. [Source:Yarbus (1967)]

10.3 VUFORIA 3D OBJECT TRACKER

Vuforia is an augmented reality platform supporting Android and iOS devices and VR glasses. Its new object tracking feature allows smart devices to recognize models and augment them in three-dimensional space. Apps can be built in Android Studio, XCode, or Unity. However, the algorithm does not yet support true 3D/volumetric targets, instead combining multiple 2D images to form a sort of cage around the object to be augmented, allowing recognition from multiple angles. This means that like suitable 2D images, target objects must have good surface contrast in greyscale, avoid repetitive patterns and be lit evenly, preferably by diffuse light with no reflections. Depending on the viewing angle, different types of objects may be ideal - with smartphones and tablets, objects are usually viewed obliquely from above, meaning that a wide, flat object with a clearly marked top are ideal, while digital eyewear requires objects that are best viewed at eye level. As the volume of the object is not yet recognized, it does not occlude the augmentation, which must be kept in mind to create realistic scenes. The workflow is simple and intuitive: an object is scanned using the standalone app and a printed target [Fig. 10.3], uploaded to the Vuforia database manager, and exported as a package appropriate to the software used for creating the app [Fig. 10.4].

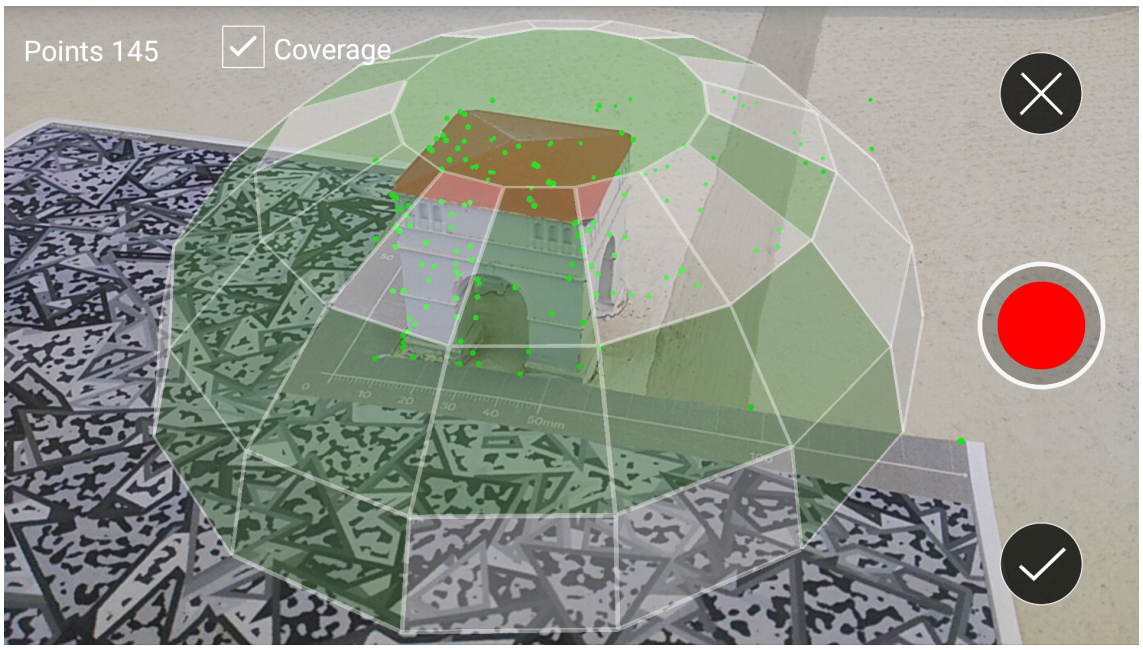


Figure 10.3: Examples of an object being scanned using a printed target.

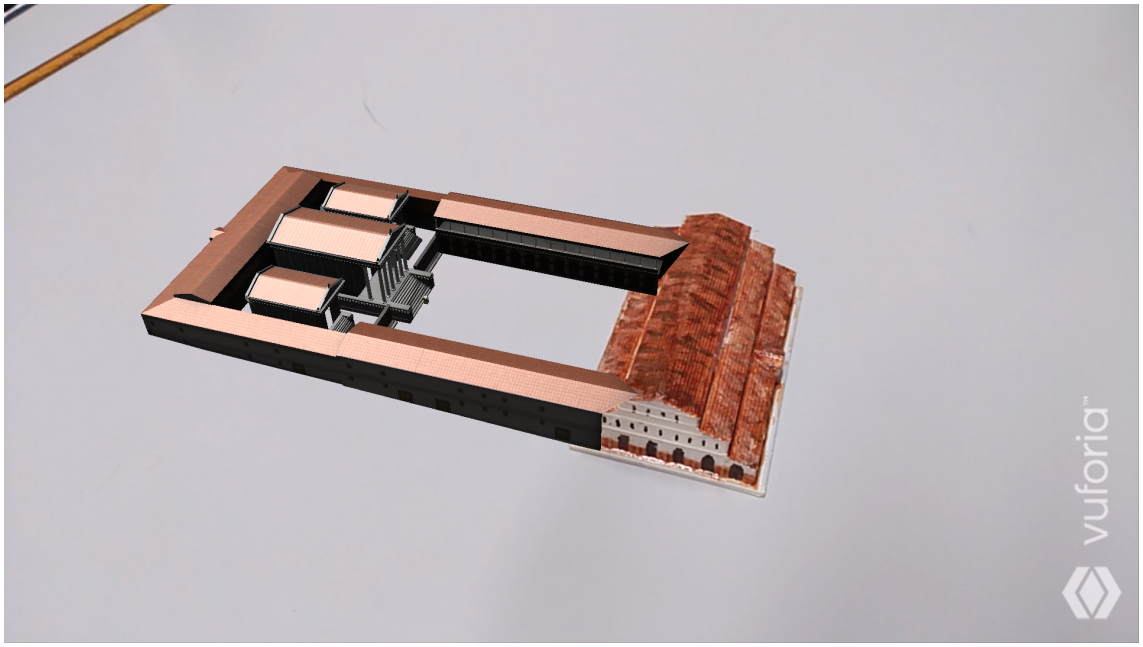


Figure 10.4: Example of Vuforia being used for the Carnuntum case study.



11 IMPACT OF ANY DEVIATION FROM THE ANNEX I

So far there are no relevant impacts of any deviation from Annex I.



12 NEXT STEPS

After the submission of deliverable 3.4 the WP3 will be completed. However, the application of object tracking methods for augmented reality applications is still in progress and the final results should be presented in later ITN-DCH events.



REFERENCES

Ali, A., & Aggarwal, J. (2001). Segmentation and recognition of continuous human activity. In *Detection and recognition of events in video, 2001. proceedings. ieee workshop on* (pp. 28–35).

Ali, S., & Shah, M. (2008). Floor fields for tracking in high density crowd scenes. In *European conference on computer vision* (pp. 1–14).

Barron, J. L., Fleet, D. J., & Beauchemin, S. S. (1994). Performance of optical flow techniques. *International journal of computer vision*, *12*(1), 43–77.

Black, M. J., & Anandan, P. (1996). The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer vision and image understanding*, *63*(1), 75–104.

Bowyer, K., Kranenburg, C., & Dougherty, S. (1999). Edge detector evaluation using empirical roc curves. In *Computer vision and pattern recognition, 1999. ieee computer society conference on.* (Vol. 1).

Breitenstein, M. D., Reichlin, F., Leibe, B., Koller-Meier, E., & Van Gool, L. (2009). Robust tracking-by-detection using a detector confidence particle filter. In *2009 ieee 12th international conference on computer vision* (pp. 1515–1522).

Brostow, G. J., & Cipolla, R. (2006). Unsupervised bayesian detection of independent motion in crowds. In *2006 ieee computer society conference on computer vision and pattern recognition (cvpr'06)* (Vol. 1, pp. 594–601).

Calandra, D. M., Di Mauro, D., D'Auria, D., & Cutugno, F. (2016). Eyecu: an emotional eye tracker for cultural heritage support. In *Empowering organizations* (pp. 161–172). Springer.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*(6), 679–698.

Cao, C., Li, C., & Sun, Y. (2015). Motion tracking in medical images. *Biomedical Image Understanding, Methods and Applications*, 229–274.

Chan, A. B., Liang, Z.-S. J., & Vasconcelos, N. (2008). Privacy preserving crowd monitoring: Counting people without people models or tracking. In *Computer vision and pattern recognition, 2008. cvpr 2008. ieee conference on* (pp. 1–7).

Comaniciu, D., Ramesh, V., & Meer, P. (2003). Kernel-based object tracking. *IEEE Transactions on pattern analysis and machine intelligence*, *25*(5), 564–577.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 ieee computer society conference on computer vision and pattern recognition (cvpr'05)* (Vol. 1, pp. 886–893).

Dehghan, A., & Kalayeh, M. M. (n.d.). Understanding crowd collectivity: A meta-tracking approach.

Dehghan, A., & Shah, M. (2016). Binary quadratic programing for online tracking of hundreds of people in extremely crowded scenes. *arXiv preprint arXiv:1603.09240*.

Dollár, P., Wojek, C., Schiele, B., & Perona, P. (2009). Pedestrian detection: A benchmark. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on* (pp. 304–311).

Gammeter, S., Gassmann, A., Bossard, L., Quack, T., & Van Gool, L. (2010). Server-side object recognition and client-side object tracking for mobile augmented reality. In *2010 ieee computer society conference on computer vision and pattern recognition-workshops* (pp. 1–8).

Guerra-Filho, G. (2005). Optical motion capture: Theory and implementation. *RITA*, *12*(2), 61–90.

Hasan, M. M., Thakur, J. M., & Podder, P. (n.d.). Face detection in color images using skin color.

Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. *Artificial intelligence*, *17*(1-3), 185–203.

Hu, W., Li, X., Luo, W., Zhang, X., Maybank, S., & Zhang, Z. (2012). Single and multiple object tracking using log-euclidean riemannian subspace and block-division appearance model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(12), 2420–2440.

Izadinia, H., Saleemi, I., Li, W., & Shah, M. (2012). 2t: Multiple people multiple parts tracker. In *European conference on computer vision* (pp. 100–114).



Jacob, R., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, *2*(3), 4.

Ji, P., Li, H., Salzmann, M., & Zhong, Y. (2016). Robust multi-body feature tracker: A segmentation-free approach. *arXiv preprint arXiv:1603.00110*.

Kim, K., & Davis, L. S. (2011). Object detection and tracking for intelligent video surveillance. In *Multimedia analysis, processing and communications* (pp. 265–288). Springer.

Kiratiratanapruk, K., & Siddhichai, S. (2006). Vehicle detection and tracking for traffic monitoring system. In *Tencon 2006-2006 ieee region 10 conference* (pp. 1–4).

Kuo, C.-H., Huang, C., & Nevatia, R. (2010). Multi-target tracking by on-line learned discriminative appearance models. In *Computer vision and pattern recognition (cvpr), 2010 ieee conference on* (pp. 685–692).

Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2015). A convolutional neural network cascade for face detection. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 5325–5334).

Lucas, B. D., Kanade, T., et al. (1981). An iterative image registration technique with an application to stereo vision. In *Ijcai* (Vol. 81, pp. 674–679).

Luo, W., Xing, J., Zhang, X., Zhao, X., & Kim, T.-K. (2014). Multiple object tracking: A literature review. *arXiv preprint arXiv:1409.7618*.

Markuš, N., Frljak, M., Pandžić, I. S., Ahlberg, J., & Forchheimer, R. (2014). Eye pupil localization with an ensemble of randomized trees. *Pattern recognition*, *47*(2), 578–587.

Massaro, D., Savazzi, F., Di Dio, C., Freedberg, D., Gallese, V., Gilli, G., & Marchetti, A. (2012). When art moves the eyes: a behavioral and eye-tracking study. *PloS one*, *7*(5), e37285.

Mitzel, D., Horbert, E., Ess, A., & Leibe, B. (2010). Multi-person tracking with sparse detection and continuous segmentation. *Computer Vision–ECCV 2010*, 397–410.

Naspetti, S., Pierdicca, R., Mandolesi, S., Paolanti, M., Frontoni, E., & Zanoli, R. (2016). Automatic analysis of eye-tracking data for augmented reality applications: A prospective outlook. In *International conference on augmented reality, virtual reality and computer graphics* (pp. 217–230).

Okuma, K., Taleghani, A., De Freitas, N., Little, J. J., & Lowe, D. G. (2004). A boosted particle filter: Multitarget detection and tracking. In *European conference on computer vision* (pp. 28–39).

Park, Y., Lepetit, V., & Woo, W. (2008). Multiple 3d object tracking for augmented reality. In *Proceedings of the 7th ieee/acm international symposium on mixed and augmented reality* (pp. 117–120).

Porikli, F., Tuzel, O., & Meer, P. (2006). Covariance tracking using model update based on lie algebra. In *2006 ieee computer society conference on computer vision and pattern recognition (cvpr'06)* (Vol. 1, pp. 728–735).

Prisacariu, V. A., & Reid, I. (2012). 3d hand tracking for human computer interaction. *Image and Vision Computing*, *30*(3), 236–250.

Quiroga, R. Q., & Pedreira, C. (2011). How do we see art: an eye-tracker study.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology*, *62*(8), 1457–1506.

Rodriguez, M., Ali, S., & Kanade, T. (2009). Tracking in unstructured crowded scenes. In *2009 ieee 12th international conference on computer vision* (pp. 1389–1396).

Rodriguez, M., Sivic, J., Laptev, I., & Audibert, J.-Y. (2011). Data-driven crowd analysis in videos. In *2011 international conference on computer vision* (pp. 1235–1242).

Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on pattern analysis and machine intelligence*, *20*(1), 23–38.



Saeed, K., Tabędzki, M., Rybnik, M., & Adamski, M. (2010). K3m: A universal algorithm for image skeletonization and a review of thinning techniques. *International Journal of Applied Mathematics and Computer Science*, *20*(2), 317–335.

Serby, D., Meier, E., & Van Gool, L. (2004). Probabilistic object tracking using multiple features. In *Pattern recognition, 2004. icpr 2004. proceedings of the 17th international conference on* (Vol. 2, pp. 184–187).

Shen, W., Zhao, K., Jiang, Y., Wang, Y., Zhang, Z., & Bai, X. (2016). Object skeleton extraction in natural images by fusing scale-associated deep side outputs. *arXiv preprint arXiv:1603.09446*.

Shi, J., & Tomasi, C. (1994). Good features to track. In *Computer vision and pattern recognition, 1994. proceedings cvpr'94., 1994 ieee computer society conference on* (pp. 593–600).

Song, B., Jeng, T.-Y., Staudt, E., & Roy-Chowdhury, A. K. (2010). A stochastic graph evolution framework for robust multi-target tracking. In *European conference on computer vision* (pp. 605–619).

Szeliski, R., & Coughlan, J. (1997). Spline-based image registration. *International Journal of Computer Vision*, *22*(3), 199–218.

Teutsch, M., & Grinberg, M. (2016). Robust detection of moving vehicles in wide area motion imagery. In *Proceedings of the ieee conference on computer vision and pattern recognition workshops* (pp. 27–35).

Tomasi, C., & Kanade, T. (1991). *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ. Pittsburgh.

Tuzel, O., Porikli, F., & Meer, P. (2006). Region covariance: A fast descriptor for detection and classification. In *European conference on computer vision* (pp. 589–600).

Uchiyama, H., & Marchand, E. (2012). Object detection and pose tracking for augmented reality: Recent approaches. In *18th korea-japan joint workshop on frontiers of computer vision (fcv)*.

Wang, Y., Georgescu, B., Chen, T., Wu, W., Wang, P., Lu, X., … Comaniciu, D. (2013). Learning-based detection and tracking in medical imaging: a probabilistic approach. In *Deformation models* (pp. 209–235). Springer.

Xu, J., Zhang, C., & Goto, S. (2011). Object tracking by detection for video surveillance systems based on modified codebook foreground detection and particle filter. In *Intelligent signal processing and communications systems (ispacs), 2011 international symposium on* (pp. 1–6).

Yang, Y., & Ramanan, D. (2011). Articulated pose estimation with flexible mixtures-of-parts. In *Computer vision and pattern recognition (cvpr), 2011 ieee conference on* (pp. 1385–1392).

Yarbus, A. L. (1967). *Eye movements during perception of complex objects*. Springer.

Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: A survey. *Acm computing surveys (CSUR)*, *38*(4), 13.

Yilmaz, A., Li, X., & Shah, M. (2004). Contour-based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Transactions on pattern analysis and machine intelligence*, *26*(11), 1531–1536.