WirelessSyncroVision: Wireless Synchronization for Industrial Stereoscopic Systems

Andry Maykol P. · A. Paulo Moreira · Paulo G. Costa

Abstract The research proposes a novel technological solution for marker-based human motion capture called WirelessSyncroVision (WSV). The WSV is formed by two main modules: the visual node (WSV-V) which is based on a stereoscopic vision system and the marker node (WSV-M) that is constituted by a 6-DOF active marker. The solution synchronizes the acquisition of images in remote muti-cameras with the ON period of the active marker. This increases the robustness of the stereoscopic system to illumination changes, which is extremely relevant for programming industrial robotic-arms using a human demonstrator (PbD).

In addition, the research presents a robust method named Adaptive and Robust Synchronization (ARS), that is designed for temporal alignment of remote devices using a wireless network. The algorithm models the phase difference as a function of time, measuring the parameters that must be known to predict the synchronization instant between the active marker and the remote cameras.

Results demonstrate that the ARS creates a balance between the real-time capability and the performance estimation of the phase difference. Therefore, this research proposes an elegant solution to synchronize image acquisition systems in real-time that is easy to implement with low operational costs however, the major advantage of the WSV is related to its high level of flexibility since it can be extended toward to other devices besides the PbD, for instance, motion capture, motion analysis and remote sensing systems.

Keywords Real-time · Remote cameras · Trigger · Programming by demonstration

Andry Maykol P.[a], Paulo G. Costa [a] and A. Paulo Moreira [a]

INESC TEC and Faculty of Engineering, University of Porto,
[a] Robotics and Intelligent Systems ,
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
Corresponding author: E-mail: andry.pinto@fe.up.pt (Andry Maykol P.), Tel.: +351 22 508 14 00, Fax: +351 22 508 14 40

1 Introduction

The robotic-arms are key elements in many manufacturing industries since they perform unsafe, hazardous, repetitive and unpleasant tasks, while working continuously with a high level of performance. Nowadays, industrial robots are programmed off-line using structured languages that require skilled programmers with a strong knowledge of the industrial setup and a detailed definition of the operation that must be conducted by each robot autonomously (its motion, interaction with other equipments and safety procedures). As expected, this is a time consuming process that requires a constant cash flow in qualified human resources which is a major disadvantage for small and medium-sized enterprises. Therefore, more intuitive methodologies for programming these robots (easier for non-expert personnel) are desired and several related works can be found in the literature [9, 8, 6, 23, 22, 6, 15].

Recently, a 6-DoF (degree of freedom) visual marker was presented in [9, 8], which is based on high intensity LEDs (Light Emitting Diodes) that turn ON during the exposer period of the cameras that are capturing the environment. The position and orientation of the active marker are retrieved through a stereoscopic vision system that makes it possible to replicate the trajectory of the wrist of the human teacher with a high degree of accuracy and effectiveness. This industrial technology[1]

---

[1] Patented by the University of Porto.

was designed for programming a robot by demonstration (PbD) and through motion imitation. Although being quite remarkable the accuracy of the imitation that is achieved by the robotic-arm, this technology has some inconveniences yet: the solution is currently limited by a single stereoscopic vision and it requires a set of cables to assure the synchronization between the optical system and the 6-DoF marker. As one can expect, these cables connect the human teacher to the multi-camera system which reduces the maneuverability of the operator and the flexibility of the system for large-scaled industrial environments.

In this context, the current research presents a novel technological solution called WirelessSyncroVision (WSV), and evaluates a designed method that synchronizes the instant of the acquisition of multi-cameras with time events of a 6-DoF active marker. In the past few years, a high number of temporal alignment methods to synchronize multi-cameras have been proposed however, the large majority focused on a software approach: unmatched bandwidth of heterogeneous network environments [12], software triggering using markers [5] and the asynchronism between multiple points-of-view [4]. These methods use elaborated techniques (shape from silhouettes and volumetric analysis) [7] but results have proved that the level of temporal alignment is often insufficient in terms of accuracy for photometric evaluations [10] and 3D reconstructions [7, 2]. In fact and by resorting to these methods, the faithful reproduction of a moving human is severely compromised for applications of PbD since noticeable texturing artifacts appear in the moving objects.

Therefore, the WSV resorts to a wireless communication for triggering remote devices at same instant of time: an external signal synchronizes the capturing of images with specific time events. Two distinct modules constitute the main elements of the WSV: the visual node (WSV-V) and the marker node (WSV-M); and the communication topology settles in a MIMO (multiple-input and multiple-output) architecture at 2.4GHz ISM (industrial, scientific and medical) band. The internal mechanism of the WSV is based on a technique that analyses the communication between the WSV-M and the WSV-V, and determines the phase delay between the electrical signal that triggers the industrial cameras (WSV-V) and the signal that activates the visual marker (WSV-M). This synchronization technique uses a sensing structure composed of distinct observation levels configured in time and defined by robust formulations, as will be discussed later. By knowing the expected delay, the WSV-V can adjust the instant of time for triggering the stereoscopic system (through electrical signals in wired cables) - clock signal - in sync

with the time period that the visual marker stays glowing. This increases the autonomy of the 6-DoF marker (the power consumption is a major concern because the WSV-M is a remote device) and increases the clock signal. In this way, the frequency for capturing the human motion can be augmented up to the physical limitations of the camera.

Therefore, the contributions of this research include:

- A novel prototype system called WirelessSyncroVision (WSV), that synchronizes several remote vision systems (for instance, stereoscopic systems) with external temporal events. The WSV was especially designed for industrial applications with real-time acquisition constraints (delays less than 20 microseconds), for instance: visual measurements, motion analysis and 3D reconstructions;
- An robust method for modeling the phase difference between multi-cameras and a visual marker, with a performance less influenced by the electromagnetic noise of industrial environments that corrupts the wireless communications. The method is called Adaptive and Robust Synchronization (ARS) and increases the robustness of the estimation process to the presence of outliers (usually, noise components);
- The WSV with this robust method increases the power efficiency of the solution and the blinking/clock frequency (sampling frequency) of the vision system which is of extremely relevance when capturing the fastest movements of the operator;
- Results of this research can directly be transfered to the industry;
- Extensive qualitative and quantitative evaluation by considering several testing industrial-like scenarios.

The experimental results include the analysis of the proposed robust technique in several contexts: a comparison to a running-average; and an evaluation and discussion of the behavior of the WSV in real and practical industrial scenarios. The performance of the WSV is evaluated using realistic environments corrupted by electromagnetic noise (robotic-arm) and by applications with limited computational resources. The quality and accuracy of the synchronization is validated using a remote stereoscopic vision system and an objective assessment metric, namely the phase difference. Experimental considerations indicate that WSV with the "robustification" of the temporal correlation between consecutive phase delays is computationally rewarding and represents an alternative to the other systems based on remote sensors/cameras. Therefore, important conclusions are obtained about the usefulness of the WSV for different type of applications. The results demonstrate that the proposed technology performs satisfac-

torily better and can be used as a tool for motion perception in other applications, namely, remote sensing, three-dimensional kinematic analysis, photogrammetry and videogrammetry.

The article is organized as follows. Section 2 presents a brief review of the approaches that are commonly used to synchronize multi-views in motion perception and 3D reconstructions. Section 3 shows the WirelessSyncroVision. An overview of the full structure of the proposed technology is provided in section 3.1 and afterwards, section 3.2 presents in detail the robust technique that is also proposed. Section 4 demonstrates the experimental results: section 4.2 shows the accuracy and robustness of the technique with the WSV for modeling the phase difference and, section 4.1 presents the industrial PbD setup that is considered case-study. Finally, section 5 presents the most important conclusions of this research.

2 Related works

Vision sensor technology is often used by industrial robotic applications because, recent technological advances have led to the appearance of low-power sensors with more sophisticated and cost-effective features, for instance, lower noise, external triggering, higher resolution and speed acquisition. Some of these features have been used by robotic-based systems, especially, for real-time visual measurements and 3D reconstruction. The triangulation's principle is the basis for the visual reconstruction of objects, and it uses the geometric correspondence of points in images and, the relative position of the optical sensors and the light of source. Therefore, synchronizing the acquisition of images between multiple sensors is crucial for ensuring that physically distributed sensors have a common notion of time. This has a significant effect on the quality of the measuring process [2], which may compromise industrial applications of programming by demonstration (PbD), the correct operation of inspecting systems in many areas like, manufacturing industry, automotive, security and surveillance systems, fault diagnosis and recovery. The goal of video synchronization is to ensure that features matched between sequences of images also correspond to the same instant in time before the spatial matching. In theory, the synchronization can be achieved through an appropriate control algorithm for the vision sensor with image-based measurements or through an external reference signal (clock triggers): software and hardware approaches, respectively.

A rich literature exists on studying algorithms (software approach) that calibrate and synchronize multiple optical sensors using geometric correspondences [19]

however, estimating the geometric parameters [11] and the time delays [7, 4] in the same procedure might sacrifice the accuracy and, therefore, it is more desirable a synchronization procedure without image correspondences [11]. Other approach exploits the epipolar tangents [18, 13]. The silhouette contours of the objects is used to calculate points where the tangent is an epipolar line [20]. Specific temporal and spatial changes of the objects' shapes are used in [17] and the research proposes a method to synchronize multiple cameras without external triggering signal. The major problem related to the image-based synchronization approaches is related to the fact that they cannot be generalized for any application or environment since they have innate constraints and assumptions: restriction of the objects movement and deterministic features.

This kind of limitation is not appropriate for industrial environments. In practice, industrial applications usually resort to optical sensors equipped with synchronization trigger inputs for shuttering timing. A hardware approach means that a signal is shared by all cameras and the signal is usually driven by a host device. This approach has several advantages related to the robustness, reliability and accuracy of the entire acquisition system: for functioning under rapid deployment conditions, operation in unstructured locations or dangerous environments. Synchronization is relatively easy to achieve in a dedicated and wired network, whose visual sensors can constantly listen for periodic events from a well-known clock server. Over the past few decades, investigators have developed clock synchronization protocols for wired networks, for instance, IEEE 1394 and Ethernet. However, the deployment of these wires is cumbersome in some industrial situations because it adds much financial cost and adds constraints on the spatial configuration of the whole perceptual system.

Other possible solution is to synchronize multiple optical sensors without a wired network. Recent researching works have proposed techniques that use external triggering signals [2] and wireless networks [3, 16]. A high number of applications in the field of remote monitoring and interaction can be designed based on wireless sensor networks and, in this context, a huge scientific effort has been made over the last years [16]: clock synchronization, energy-efficiency, hardware and bandwidth limitations, unstable connections, latency, loss of messages and network dynamics [21]. Most of these researches are unsuitable for real-time robotic applications based on wireless optical sensors, especially, in industrial environments because the challenges posed by wireless networks are theoretically different [11, 3, 16] and the proposed methods have difficulty in assur-

ing the real-time requirements that are needed by these systems.

External devices can be used for remote triggering of the cameras. Ricardo Barros et. al (2006) [2] presented a scheme that synchronizes multiple cameras by transmitting a coded audio signal over radio frequency and for receivers that are connected to the microphone input of the visual sensors. In this work, the authors have showed that the phase difference between cameras is measured with a resolution of a millisecond and it changes linearly over time. Andry Pinto (2012) [1] proposed infrared signal protocol modulated in frequency that makes possible the synchronization of remote moving cameras. This approach is interesting however, infrared signals may not be suitable for industrial environments. Finally, a multi-camera system for marker-based human motion capture is presented in [5], which exploits the epipolar geometry and the particle likelihood (annealed particle filter) evaluation to avoid noisy and false measurements of the markers position. The system achieved an accuracy in the range of centimeters which is usually not enough for PbD applications in industrial scenarios.

This paper shows that hardware approaches resorts to external devices based on the wireless synchronization of visual sensors is possible and it offers a higher performance in terms of flexibility, reliability and robustness, when compared to software approaches (that eventually lead to a lower financial cost). In addition, it describes a solution that assures a real-time synchronization (in range of a few microseconds) and, can be deployed in several scenarios (motion analysis and 3D photogrammetry) under different environmental conditions (domestic or industrial).

3 WirelessSyncroVision

The blinking of high intensity RGB (Red, Green and Blue) LEDs synchronized with the shuttering of a camera belonging to a stereoscopic system makes the visual acquisition highly immune to external light sources [1]. These LEDs emit a high intensity light overshadowing most of the other light sources and thus, the color detection is more robust to changes in brightness. Other feature of this approach is that the LEDs are switched-on during a very short period of time which does not affect people's vision, decreases the energy consumption and increases the longevity of the LEDs [1]. This concept is called SyncroVision (patented) and it is being developed at the Faculty of Engineering of the University of Porto.
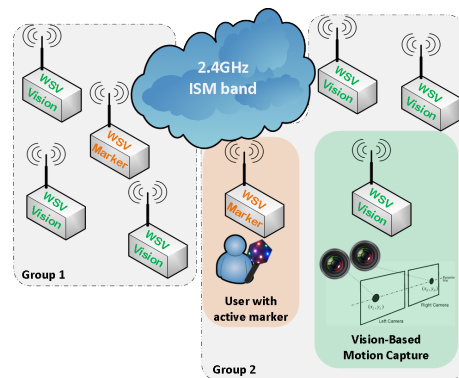
A 6-DoF marker with several LEDs at standard positions resorts to this technology and additional infor-

mation about the marker can be found in [9] . The conventional SyncroVision requires an accurate synchronization between the marker and the shuttering of the camera. This decreases the influence of the ambient lighting but the temporal alignment is usually done at the expense of wired cables. In most of the cases, wired cables are desired for industrial environments since they are substantially immune to electromagnetic interferences however, for applications of programming by demonstration (PbD), the cables that connect the human operator to the multi-camera system affect the maneuverability of the operator which compromises the phase of teaching the robotic-arm (operator's motion). Moreover, wired cables reduce the flexibility and the operating range of the SyncroVision in large-scaled industrial environments.

Therefore, this research focused in designing and developing a wireless approach for synchronizing the active marker with remote stereoscopic systems. Besides the programming by demonstration, this research can be directly applied to other application areas namely: 3D photogrammetric analysis, motion detection and fusion of multi-sensor data.
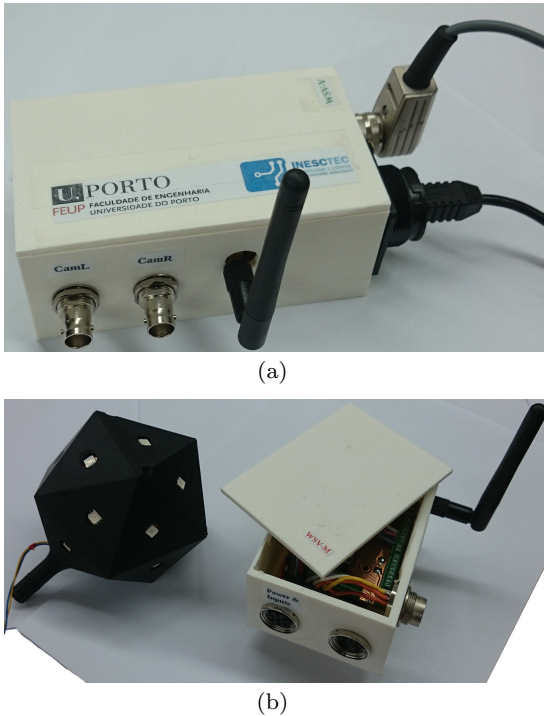
3.1 The concept

The proposed wireless system is called WirelessSyncroVision and it is formed by two main modules: the visual node (WSV-V) and the marker node (WSV-M). The visual node (WSV-V) of figure 2(a) is based on a wireless module, triggering system and two or more cameras with external trigger. The major goal of WSV-V is to grab frames by sending electrical signals to the external trigger when it receives pre-defined messages with a unique and distinctive identification number. At the other side, the marker node (WSV-M) of figure 2(b) has the wireless module and the triggering system for the 6-DoF marker.

Fig. 1 The concept of the WirelessSyncroVision for an industrial scenario of PbD.

The synchronization between WSV-M and WSV-V is accomplished by a wireless communication at 2.4GHz ISM (industrial, scientific and medical) band, see figure 1: an initial message is sent by the WSV-M to the WSV-V. The wireless triggering signals are driven by marker nodes (WSV-M) and all signals can be shared by all visual nodes (WSV-V) in the environment. If the received message has the correct identification number, then the WSV-V node triggers the multi-camera system (acquiring the images) and sends an acknowledge message back to the WSV-M. At the same time, the WSV-M estimates the network delay based on a robust method that analysis the network (presented in section 3.2) and lights the 6-DoF marker (activating the set of LEDs) after the estimated time delay. In this way, the markers remains ON during the shuttering period of the multi-cameras which reduces the power consumption, increases the accuracy (real-time) of the synchronization process and enables high speed acquisition - high frame rates.
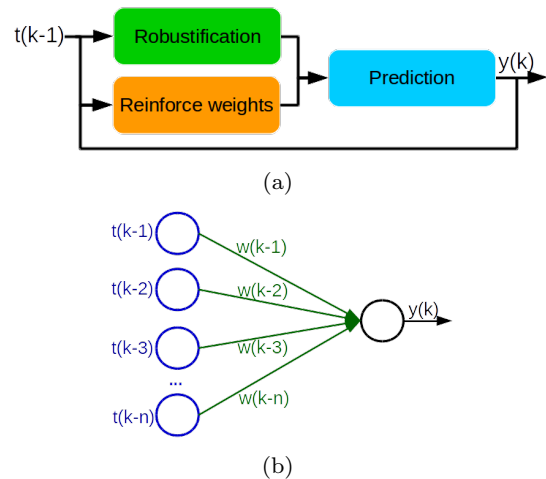

(a)

(b)

Fig. 2 Examples of the modules designed for the WirelessSyncroVision: the WSV-V (on left) is the visual module that triggers multi-cameras and the WSV-M (on right) depicts the operator's module with the 6-DoF marker (represented in black).

3.2 Modeling the phase difference between multi-cameras and a remote visual marker

The phase difference can be estimated by measuring the communication delays over time. Each observation, $t_k = t(k)$, measures the time that comprises the elapsed period where the WSV-M sends a wireless signal and receives an acknowledge from the WSV-V[2], and it represents the $2\times$ TOA (time of arrival). However, this process may suffer from the presence of outliers originated by loss messages (causing timeouts) and noise of the industrial environment that corrupts the communications. Consequently, the synchronization instant based on the previous observation (or a running average) may be poorly estimated and unreliable because the formulation is incapable of distinguishing between inlier and outlier observations.

Therefore, this section presents a robust method for temporal alignment of two remote devices using a wireless network called, Adaptive and Robust Synchronization (ARS). The method resorts to robust statistical techniques to define a function of the temporal domain based on sequential observations (phase differences of the WSV-M and WSV-V). The ARS implements an online regression technique with an adaptive mechanism that predicts the expected TOA (TOA = $y_k/2$). An overview schematic of the method is depicted in figure 3(a) and, as can be noticed, it is composed by three stages: "robustification", reinforce weights and prediction.

(a)

(b)

Fig. 3 Schematic of the Adaptive and Robust Synchronization (ARS) method which is formed by three stages: "robustification", prediction and reinforce weights. Figure 3(b) demonstrates a diagram of prediction phase.

---

[2] In scenarios having multiple WSV-V modules, the observation that is considered by the algorithm will be the one with the lowest value.

Describing each phase in more detail and beginning by the robustification stage, the goal of a regression technique is to resolve the model parameters that produce the best fit for a region of n independent observations (t_k) [14]. This means that an objective function can be expressed using M-Estimators (*Maximum Likelihood Type Estimators*) and the residuals $err_k = t_k - y_k$.

$$min \; \xi = \sum_{i=1}^{n} \varphi(t_{k-i} - y_{k-i}); \qquad (1)$$

where t_k is the phase difference and y_k is the estimated phase difference at time instant k.

A detailed description of the robust theory with application to filtering image sequences can be found in [14]. A robust function $\varphi(err_k)$ decreases the influence of outliers by replacing the squared residuals by a robust error function, for instance, the *Charbonnier*. The term "robust" is conventionally related to the rejection of outliers by recovering the estimate that represents the majority of the data [14]. A influence function $\psi(err_k)$ is the derivative of $\varphi(err_k)$ and, it is used to solve equation 1. This function determines the influence level that a particular observation has on the solution. A final function can be obtained by the weight function, $\omega(err_i)$ which is defined by:

$$\omega(t_i - y_i) = \frac{\psi(t_i - y_i)}{t_i - y_i}. \qquad (2)$$

This defines a weight contribution according to the residual value (measures similarity level of the associated observation at different temporal instants). Therefore, the robustification phase receives the t_k and determines the weight contribution w_{k-1} of the previous y_k.

The prediction stage recovers the estimate that represents the majority of the observations, by returning a reliable estimative of the y_k and at the current instant of time, see Equation 3.

$$y_k = \frac{\sum_{i=1}^{n} w_{k-i} . t_{k-i}}{\sum_{i=1}^{n} w_{k-i}}. \qquad (3)$$

where $w_i = \frac{1}{\sqrt{1 + \frac{err_i^2}{\beta^2}}}$ and β is a setting parameter of the *Charbonnier* estimator. In Equation 3, it is evident that the value of y_k is obtained by weighting the contributions of all temporal neighbors however, the weights make possible to detect and reject the influence of outliers.

Finally, the ARS introduces a mechanism that updated the weights according to the estimated y_k and the observed t_k. This procedure is the final stage of the process depicted in figure 3(a) and it is conducted when the acknowledge message of the WSV-V arrives to the WSV-M: the t_k is available. This stage analyzes if the predicted value of y_k was reliably estimated and measures the confidence level that updates the entire process, see Equation 4.

$$w_{k-i} \longleftarrow w_{k-i} + \rho(t_{k-i} - y_k), \forall i \in \{1, ..., n\}. \qquad (4)$$

where ρ is a learning rate that limits the influence of undesirable outliers in the process of updating the weights. The main objective of Equation 4 is to enable the evolution of weights over time, making possible small adjustments for a better fitting of the contribution of each temporal neighbor to the current state of the synchronization process.

Thus, the adaptive structure of the ARS is continuously updating the neighbors and reinforcing the weights that balances the contribution of each observation to the next estimative of y_k. Besides their simplicity, the proposed algorithm have several advantages: a real-time and robust prediction of the phase difference, casual loss messages have a minor effect in the estimation and an online mechanism for updating the contribution of each neighbor (w_i) in accordance to the evolution of the network conditions (availability over time).

4 Experiments and Results

An extensive set of experiments was conducted as part of this research. The experiments aimed at evaluating and analyzing the behavior of the proposed WirelessSyncroVision (WSV), especially, the accuracy on estimating the instant of time that synchronizes the modules WSV-M and WSV-V, at different testing conditions.
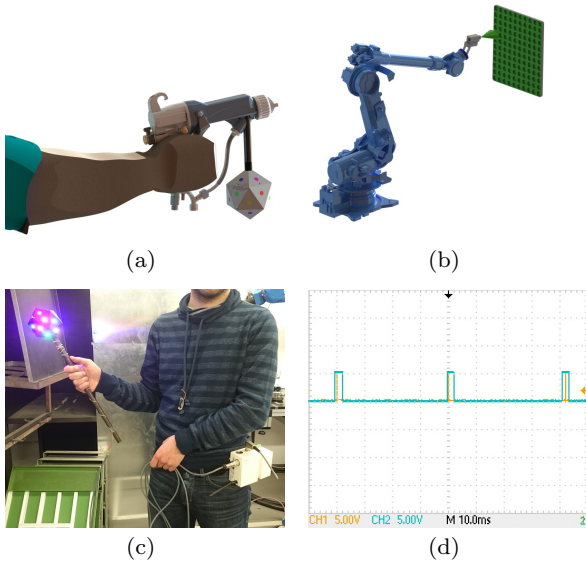
In these experiments, the estimative obtained by the ARS[3], see section 3.2, is compared to the real phase difference of several evaluation scenarios: different distances, the environment being corrupted with electromagnetic noise (of an industrial robotic-arm) and, the impact of the learning mechanism is also studied. The performance of the WSV is evaluated in the PbD scenario presented in section 4.1. Therefore, these trials capture different working conditions and possible configurations for the ARS method in the WSV.

---

[3] The method was implemented in the programming language C and for a micro-controller ATmega32u4.

4.1 Programming by demonstration

The synchronization of multi-sensors is important in faithfully reproducing the movement of humans. This is of extremely importance on industrial applications of programming by demonstration (PbD), where a robotic-arm learns how to move or grab a part by replicating a human behavior. This section shows the WSV in a programming by demonstration scenario where a human operator with the 6-DOF marker (in the WSV-M) is teaching an industrial robotic-arm how to paint special parts, see figures 4(a) and 4(b). A stereoscopic visual system is capturing the path (position and orientation) that is performed by the active marker to enable a reliable imitation that will be conducted by the robot.

(a)   (b)

(c)   (d)

Fig. 4 Figures 4(a) and the 4(b) show the concept of the PbD in a scenario where a human operator teaches a robotic manipulator how to paint a specific part. Figures 4(c) depicts the a real scenario where the human operator is connected to the WSV-M. Figure 4(d) demonstrates that the trigger of the cameras (orange line) is in sync with the ON period of the active marker (green line).

Figure 4(c) presents the industrial setup for the PbD scenario that is considered the case-study of this research[4]. As can be noticed, the 6-DOF active marker [9] within the WSV-M is synchronized with the stereoscopic vision in the WSV-V, see figure 4(d). Therefore, the WSV makes it possible to analyze the behavior of an operator by extracting the position and orientation of the active marker (by detecting the color and posi-

---

[4] A video can be found in `http://paginas.fe.up.pt/~dee10015/_research.htm`

tion of each LED). This information is required for the imitation phase that is conducted by the robotic-arm.

4.2 Robust modeling of the phase difference

Table 1 demonstrates the results in terms of $err_k = t_k - y_k$ that were obtained by the ARS (a instant of time was already presented in figure 4(d)). The evaluation was conducted in different testing conditions namey: A - close range (0.15 meters) and open space; B - medium range (8.5 meters) and open space; C - medium range (8.5 meters) and close space (non-line-of-sight since because two glass walls were separating the WSV-M from the WSV-V); and D - close range (3.5 meters), close space (non-line-of-sight since the WSV-V was occluded with a plastic enclosure) and one robotic-arm working. The ARS is compared with a running-average using the previous 5 observations. In addition, the influence of the mechanism that reinforces the weights is initially evaluated by comparing the performance of the proposed method in two configurations: $\rho = 0.01$ and $\rho = 0$; however, its influence is studied in more detail in graphs 5(a) to 7(b).

It is evident in Table 1 that the ARS obtained the highest performance due to its lowest error in all experiments. The mean error of the running-average in scenario "A" was about 3.4 microseconds (us) and, the err_k of the ARS with $\rho = 0.01$ and $\rho = 0$ was 1.3us and 2.1us, respectively. An interesting outcome can be observed in these testing conditions: the number of time-outs is significantly high due to the small distance between WSV-V and WSV-M, and because the LNA (low-noise amplifier) of the receiver is being saturated by the power amplifier (PA) of the emitter since both modules have transceivers (nRF24L01) that were configured to 2Mbs of air data rate which increases their sensitivity to noise.
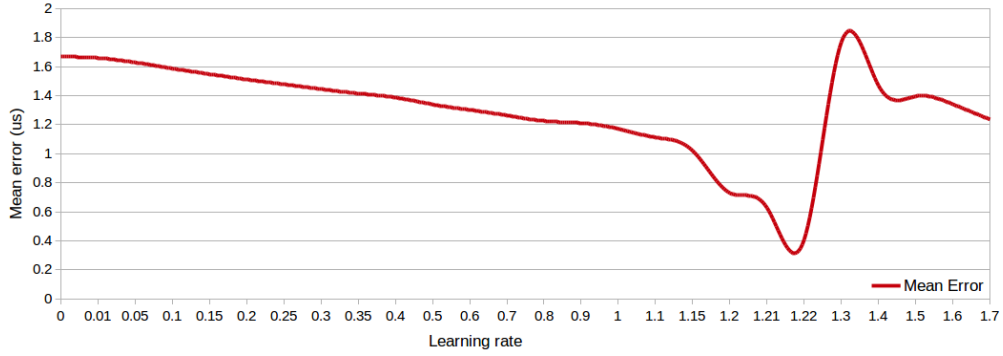
Overall, the stage that reinforces the weights is enhancing the y_k estimative because it enables small adjustments to the contribution of each neighbor and makes possible to the ARS be adaptive to changes in the working conditions. A justification to this statement is depicted in "D"-trial since the error of ARS($\rho = 0.01$) is smaller than ARS($\rho = 0$) however, both errors are very similar. This identical performance can be justified by the presence of the electromagnetic noise that corrupts the communication structure: causing a higher fluctuation of the TOA that is being observed by the WSV-M. Moreover, the ARS($\rho = 0.01$) led to a standard deviation that is small when compared to the ARS($\rho = 0$) or even the running-average.

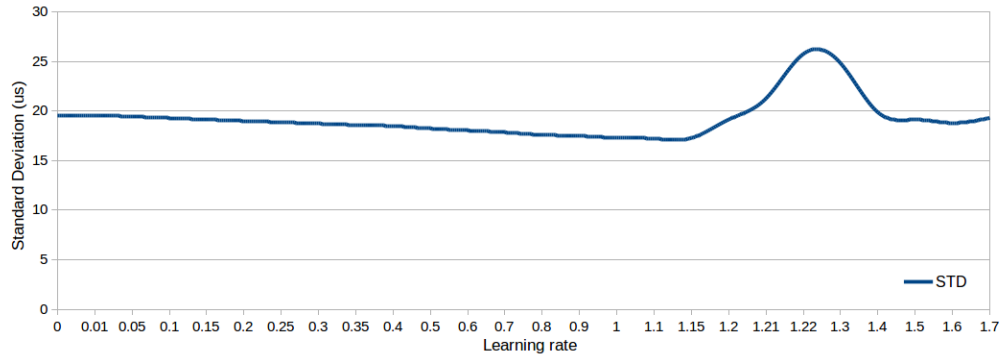The following results are based on the testing scenario "C" and they demonstrate the influence that each

Table 1 Error of predicting the phase difference (err_k). The values for the mean and standard deviation (STD) of 15000 samples are represented in microseconds. The number of timeouts is each trial is also presented and represents an average of 20 experiences conducted in the same conditions.

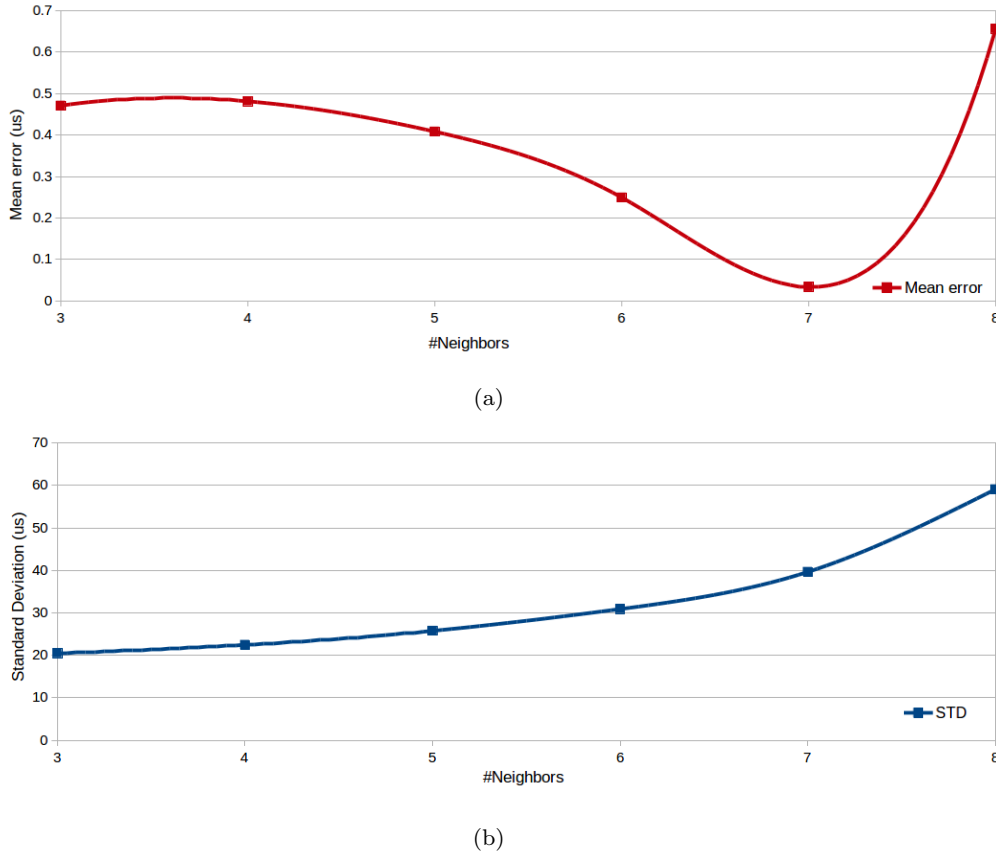| Testing scenario	$\rho = 0.01$		$\rho = 0$		Running Average		#Timeouts
	Mean	STD	Mean	STD	Mean	STD	
A (close range)	**1.280**	15.999	2.050	26.884	3.407	26.979	90
B (medium range)	**1.298**	15.231	1.434	16.397	2.351	16.423	13
C (close space)	**1.625**	15.544	1.673	18.537	2.859	18.564	3
D (noise)	**1.432**	16.724	1.472	19.954	2.262	20.280	9

(a)

(b)

Fig. 5 Figures 5(a) and 5(b) present the mean and standard deviation of err_k as a function of the learning rate, that were obtained by the ARS in "C" scenario with $\beta = 50.0$ and 5 temporal neighbors.

setting parameter (ρ, β and number of neighbors) has on the behavior of the ARS. First, the figure 5(a) represents the prediction error as a function of the learning rate. The results of this graph shows that the best performance ($err_k = 0.32$ us) is obtained for $\rho = 1.22$. However, the standard deviation is high and close to 27 us. In this way, a good ρ is by convenience the value of 1.15 since it has an acceptable error and the smallest standard deviation.

Figures 6(a) and 6(b) presents the err_k with a different number of neighbors and considering the $\rho = 1.22$ and $\beta = 50.0$. The ARS obtained the best performance when 7 temporal neighbors are used to predict the value of y_k and, under these conditions, the error was about

0.3 us. An interesting result is related to the standard deviation that increases (almost linearly) by changing the number of neighbors, see figure 6(b). Figures 7(a) and 7(b) depict the influence that the parameter β (of the Charbonnier M-Estimator) has on the ARS. Both figures are similar to the graphs of the learning rate, which means that the influence of β is similar to the ρ. In scenario "C" ($\rho = 1.22$ and 7 temporal neighbors contributing to the estimative y_k), the proposed method achieved a mean error close to 0.1us when $\beta = 50$ however, the process led to a high standard deviation (42 us). An acceptable β is perhaps the value of 55 since the ARS method obtained an err_k close to 1us and a standard deviation of 17us, which is the

(a)

(b)

Fig. 6 Figures 6(a) and 6(b) present the mean and standard deviation of err_k as a function of the number of neighbors, that were obtained by the ARS in "C" scenario with $\rho = 1.22$ and $\beta = 50.0$.
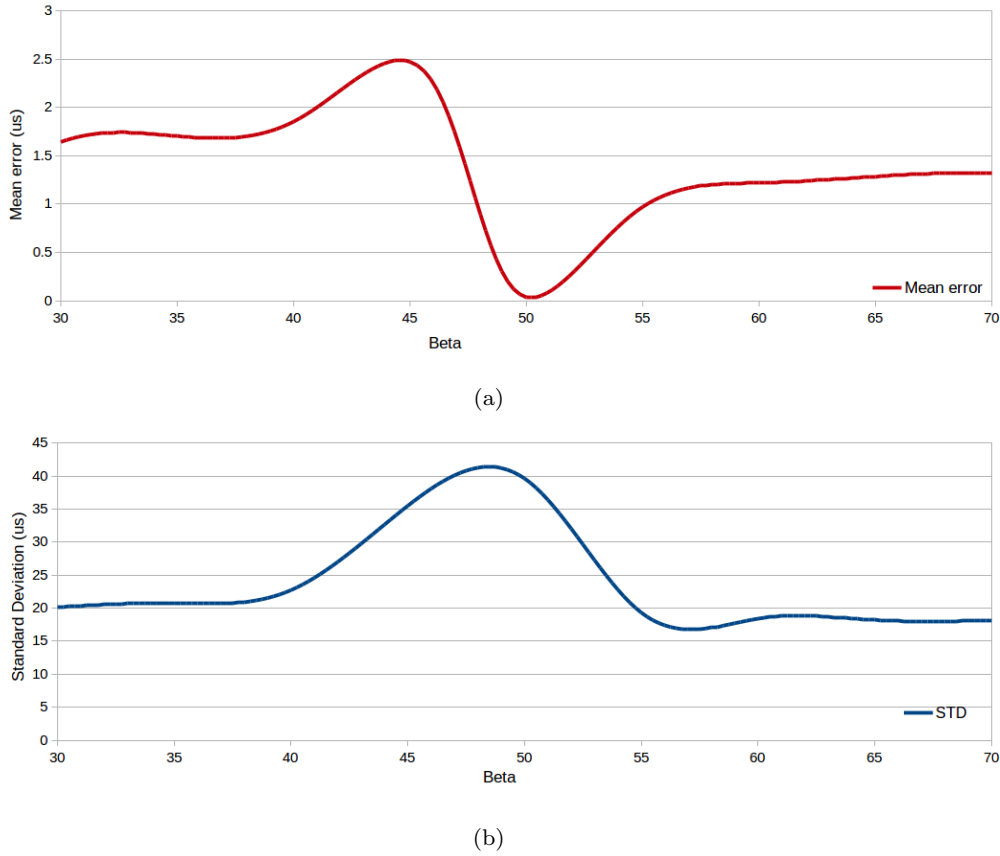
lowest value in the graph. From the previous results, a final remark must be made in accordance to the importance of the standard deviation in the WSV, because even with a configuration that made possible a small prediction error err_k, these parameters also led, unfortunately, to a high value of standard deviation in the majority of cases. This means a low confidence for the y_k and, therefore, the ON period of LEDs in the WSV-M can be unsynchronized by a little with the shuttering of cameras in the WSV-V. Although this issue be undesired, the ARS method achieves a synchronization accuracy of 20us, during the worst case scenario and contemplating the influence of the standard deviation.

Moreover, the previous results have demonstrated how the setup of the ARS can be made. In addition, they proved the ability of the method to predict the phase difference with a high accuracy level between the WSV-M and WSV-M. However, the best accuracy has associated a high value for the standard deviation. Usually, a high standard deviation means a low confidence level for the predicted value of y_k. In this way, the parametrization of the method should be balanced

in terms of accuracy and standard deviation that can compromise the synchronization procedure.

5 Conclusions

Modeling dynamic objects using multi-cameras placed in the environment requires high synchronization level for capturing different points-of-view and to make possible the reliable reconstruction of the position and the form of objects at particular instants of time. Existing systems that synchronize multi-cameras are usually inefficient (software-based), expensive (hardware-based) and do not achieve a real-time or a high level of accuracy that are required for industrial applications of PbD (programming-by-demonstrator). In this context, this article proposes the WirelessSyncroVision (WSV) which is a novel technological solution formed by the visual node and the marker node. Moreover, a robust method for synchronizing the instant of acquisition of multi-cameras with a 6-DoF active marker is also presented. This method is called Adaptive and Robust Synchronization (ARS) and models the phase difference as a function of time by predicting the parameters that

(a)

(b)

Fig. 7 Figures 6(a) and 6(b) present the mean and standard deviation of err_k as a function of the β (Charbonnier parameter), that were obtained by the ARS in the "C" scenario with 7 temporal neighbors and $\rho = 1.22$.

must be known for the temporal alignment between the active marker and the remote cameras.

The experiments conducted proved that the proposed technological solution meets the real-time requirements of common robotic systems, since it estimated the temporal delay with a mean error that was lower than 20 microseconds in the worst case scenario and, contemplating electromagnetic noise, medium distances and the enclosure of modules. In addition, incorporating a mechanism that reinforces weights of the temporal neighbors in the ARS is advantageous. Comparatively to the running-average, the ARS method achieved the best accuracy in predicting the phase difference with a good computational efficiency (runs on a microcontroller).

In short, a reliable and efficient solution is proposed as part of this study that makes it possible to synchronize remote visual acquisition sensors with time events (a active marker blinking at high frequencies). Besides the industrial application depicted, the proposing system can be used for other applications, namely, remote sensing, three-dimensional kinematic analysis, photogrammetry and videogrammetry.

References

1. A. Paulo Moreira Andry Maykol G. Pinto and Paulo G. Costa. Indoor localization system based on artificial landmarks and monocular vision. *TELKOMNIKA*, 10(4):1609–620, 2012.

2. Ricardo Barros, Tiago Russomanno, Rene Brenzikofer, and Pascual Figuerosa. A method to synchronise video cameras using the audio band. *Journal of Biomechanics*, 39(1):776–780, 2006.

3. Milan Bradonjic, Eddie Kohler, and Rafail Ostrovsky. Near-optimal radio use for wireless network synchronization. *Theoretical Computer Science*, 45(3):14–28, 2012.

4. D.N. Brito, F. Padua, R. Carceroni, and G. Pereira. Synchronizing video cameras with non-overlapping fields of view. In *Brazilian Symposium on Computer Graphics and Image Processing*, pages 37–44, 2008.

5. C. Canton-Ferrer, J. Casas, and M. Pardas. Towards a low cost multi-camera marker based human motion capture system. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 2581–2584, Nov 2009.

6. R. Dillmann, O. Rogalla, M. Ehrenmann, R Zollner, and M. Bordegoni. Learning robot behaviour and skills based on human demonstration and advice: the machine learning paradigm. In *International Symposium In Robotics Research*, volume 9, pages 229–238, 2000.

7. Tobias Duckworth and David J. Roberts. Camera image synchronization in multiple camera real-time 3d reconstruction of moving humans. In *IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications*, volume 1, pages 138–144, 2011.

8. M. Ferreira, P. Costa, L. Rocha, and A Paulo Moreira. Stereo-based real-time 6-dof work tool tracking for robot programing by demonstration. *The International Journal of Advanced Manufacturing Technology*, 1(13):604–616, 2014.

9. M. Ferreira, P. Costa, L. Rocha, A. Paulo Moreira, and Norberto Pires. New marker for real-time industrial robot programming by motion imitation. In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 1, pages 742–747, 2014.

10. Ayman Habib, Ivan Detchev, and Eunju Kwak. Stability analysis for a multi-camera photogrammetric system. *Sensors*, 14(8):15084–15112, 2014.

11. Lei Hou, Shingo Kagami, and Koichi Hashimoto. Frame synchronization of high-speed vision sensors with respect to temporally encoded illumination in highly dynamic environments. *Sensors*, 13(4):4102–4121, 2013.

12. Jenq-Shiou Leu, Wei-Hsiang Lin, Hung-Jie Tzeng, Chi-Feng Chen, and Mu-Sheng Lin. Adaptive frame synchronization for surveillance system across a heterogeneous network. *Engineering Applications of Artificial Intelligence*, 25(7):1349 – 1354, 2012.

13. P. Mendonca, K. Wong, and R. Cipolla. Epipolar geometry from profiles under circular motion. *IEEE Trans. Patt. Anal. Mach. Intell.*, 23(1):604–616, 2001.

14. Andry Maykol Pinto, Paulo G. Costa, Miguel V. Correia, and A. Paulo Moreira. Enhancing dynamic videos for surveillance and robotic applications: The robust bilateral and temporal filter. *Signal Processing: Image Communication*, 29(1):80–95, 2014.

15. David Samper, Jorge Santolaria, Francisco Javier Brosed, and Juan Jos Aguilar. A stereo-vision system to automate the manufacture of a semi-trailer chassis. *The International Journal of Advanced Manufacturing Technology*, 67(9-12):2283–2292, 2013.

16. Bharath Sundararaman, Ugo Buy, and Ajay Kshemkalyani. Clock synchronization for wireless sensor networks: a survey. *Ad Hoc Networks*, 3(1):281–323, 2005.

17. M. Ushizaki, T. Okatani, and k. Deguchi. Video synchronization based on co-occurrence of appearance changes in video sequences. In *International Conference on Pattern Recognition*, volume 3, pages 171–74, 2006.

18. B. Vijayakumar, D. Kriegman, and J. Ponce. Structure and motion of curved 3d objects from monocular silhouettes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 327–334, 1996.

19. A. Whitehead, R. Laganiere, and P. Bose. Temporal synchronization of video sequences in theory and in practice. In *IEEE Workshop on Motion and Video Computing*, volume 1, page 132137, 2005.

20. J. Yan and M. Pollefeys. Video synchronization via space-time interest point distribution. In *Advanced Concepts for Intelligent Vision Systems*, volume 1, pages 12–21, 2004.

21. Cuicui Zhang, Xuefeng Liang, and Takashi Matsuyama. Generic learning-based ensemble framework for small sample size face recognition in multi-camera networks. *Sensors*, 14(12):23509–23538, 2014.

22. R. Zollner, O. Rogalla, and R. Dillmann. Integration of tactile sensors in a programming by demonstration system. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 2, pages 1114–1119, 2002.

23. R. Zollner, O. Rogalla, and R. Dillmann. Integration of tactile sensors in a programming by demonstration system. In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 3, pages 2578–2583, 2014.