DOI: 10.2352/ISSN.2470-1173.2016.4.ERVR-422

Visual Knowledge feedback through multi-camera system aided by Augmented Reality

Josué-Rafael Montes [*a], ***Teresa Hernández-Díaz*** [*a], ***Alberto Vázquez-Cervantes*** [*a], ***Juan-M García-Huerta*** [*a], ***Leonardo Barriga-Rodríguez*** [*a], ***JA Soto-Cajiga*** [*a], ***Hugo Jiménez-Hernández*** [*a]

[*a] ***Centro de Ingeniería y Desarrollo Industrial, Av. Pie de la Custa No. 702. Desarrollo San Pablo C. P. 76130. Querétaro, Qro. México.***

Abstract

Multi camera vision system has been adopted as a versatile tool for monitoring continuously dynamics environment. Along of time several authors have worked with approaches such that offers high level interpretations over those environments. However, each one of these approaches are founded in specific theoretical sustenance. These diversities may generate heterogeneity of results; which it makes them incomparable. In order to propose a useful way to use the knowledge, it is necessary to develop approaches to unify and query these data. This work proposes an approach to unify heterogeneous data, establishing several metrics criteria for visually query criterion aided by augmented reality.

The proposal describes three principal stages, a) Catch the dynamic of movement by several approaches; b) A Similarity criterion to query by example spatio-temporal information reference and c) An Overlapping criterion of data repository aided by Augmented Reality.

The experimental model presents a new way to query information in heterogeneous data bases obtained by different approaches, finding similar trajectories using visual query-by-example approach and displaying the results aided by Augmented Reality infrastructure.

1.- Introduction

One area that has had more impact in recent years in the Artificial Vision and Digital Image Processing is in Monitoring and Multi-camera surveillance system. Actually, the Monitoring and Multi-camera surveillance system has a very important role due to among other things help to get dynamic motion of interest objects in an efficient way.

The dynamics of movement generate by the objects of interest along time is constituted by different models M_i of different nature. Thus, for a Multi-camera system with a global flow reference t_i, the dynamics is defined as a 2 tuple as follows (t_i, Φ_i); where $\Phi \in P(M), M = \{M_1, M_2, ..., M_k\}$ and $P(M)$ is the power set of M; in other words, it is the sequence of events explained by a set of models for each instant t_i.

For each $M_i \in M$, it has $M_i = \{\Psi_I, I_i, O_i\}$, such that Ψ it is a Turing Machine that it decides over the input I_i, getting as result the output O_i. Note that the nature of Ψ_i was defined by the nature of the program encoded on it. Further, for each one time sample $\{t_i, \Phi\}$, the output set of information is defined as $\Theta_i = O_i^j \mid (\Psi_i^j, I_i^j, O_i^j) \in \Phi$, which it is consequence of the amount of process that explain the sample t_i.

An order relation R is defined over Θ_i. This order relation defines superposition criteria for a particular time t_i and establish an hierarchical structure over Φ.

Next, a temporal segment of information is defined as: $S = (\Theta_1, \Theta_2, ..., \Theta_k)$; such that it represents a time-sequences of Θ_i, with a given length of k. After, for a given temporal segments S_1, S_2 with a length of k_1 and k_2 respectively. A similarity criterion $d(S_1, S_2)$ is defined. This criterion is used as a way make queries spatially and temporally [4] [10].

Finally, the results are presented in augmented reality device, superimposing and mixing the total of information of the scenario.

An experimental model is defined for testing our proposal; such that M is built with follows basic models:

- 1. An hierarchical model based on rules, which it makes an approach to express the movement as a set of symbols, and a formal grammar to recognize which sequences are allowed; thus allowed trajectories are recognized by the grammar $G = \{\sum, S_0, V, R\}$ [1].
- 2. Temporal differences; which it denotes the dynamic as the most likely zones with stimuli of movement. This model uses the differences over time to denote motion zones and the stimuli is denoted as a decay function [2].

Above models, provided temporal information of the dynamics of an environment over a set of fixed cameras. Finally, this proposal provide an adaptable framework for storing and managing the information, which they will be displayed in augmented reality devices.

2.- Topological Spaces

In order to define a similarity criterion between temporal data templates S_1 and S_2 obtained from $P(M)$; topological spaces are taken into account. In this manner, $d(S_1, S_2)$ is particularly defined as a distance on the topological space of Hausdorff as the similarity criterion.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.4.ERVR-422

Topological spaces are no more that a set whit subsets (called open sets) which satisfy three properties:

Let X be set and $\mathscr{U}$ a collection of subsets of X, then:

- (i) $\emptyset \in \mathscr{U}, X \in \mathscr{U}$.
- (ii) the intersection of two elements $\mathscr{U}$ is in $\mathscr{U}$
- (iii) The union of every set of elements of $\mathscr{U}$ is in $\mathscr{U}$.

If a topological space is metrizable then for each pair of different points $x, y \in X$ there exist open sets U_x y U_y, which contain x and y respectively, such that $U_x \cap U_y = \emptyset$. The proof is straightforward: since $x \neq y$, $d(x,y) = 2\varepsilon$ for some ε, where d is a metric of X, which give us the structure of topological space in X. Sets $B_\varepsilon(x) = z \in X; d(x,z) < \varepsilon$ and $B_\varepsilon(y)$ then meet the required conditions.

2.1 Definition

A topological space is considered as Hausdorff if for each pair of different points $x, y \in X$ there exist open sets U_x y U_y, which contain x and y respectively, such that $U_x \cap U_y = \emptyset$.

In this manner every metrizable space is considered as Hausdorff, particularly R^n with usual topology and any space with the discrete topology are Hausdorff [3].

From now on the topological space of Hausdorff will be referred as Hausdorff distance, since it is used to find similarities between trajectories in $P(M)$. Performing the comparison of trajectories $S_1, S_2, ..., S_n$ in temporal queries of M.

2.2 Hausdorff Distance

In Artificial vision and pattern recognition remains the need to define a measure to estimate the degree of similitude between different approaches that determine behavior in a scenario of one or more targets. This need is even greater considering the diversity of approaches that are generating heterogeneity of information on the results.

In the actual work, the diversity of heterogeneous information represents i.e. a trajectory performed by one target obtained by different approaches. In order to find the similitude between every one of them there must be features that represent it.

Thus, trajectories $S_1, S_2, ..., S_n \in P(M)$ can be represented as unidimensional vectors of the image coordinates values (x, y), which are named features. A comparison to extract similitudes of features is precisely the Hausdorff distance.

Hausdorff distance of two sets A and B is calculated by:

$$h(A,B) = max_{b \in B}(min_{a \in A} d(A,B)) \quad (1)$$

$$d(A,B) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + ... + (a_n - b_n)^2} \quad (2)$$

where A and B are vectors that contain $\{a_1, a_2, ..., a_n\}$ and $\{b_1, b_2, ..., b_n\}$ respectively, $d(A,B)$ is a distance (usually the Euclidean) expressed in 2, expressions $max(x)$ and $min(x)$ are functions which calculate the maximal and minimal values correspondingly.

Equation 1 is used as follows. Suppose an A vector with length of k and a B vector of length m, such that $A = \{a_1, a_2, ..., a_k\}$ and $B = \{b_1, b_2, ..., b_m\}$. Then the first element of $A; a_1$ is taken to calculate the euclidean distance of it and every element of B. This way, there are m distances between a_1 and the elements of B. From this set of distances the minimum is selected. The same process is performed for every element of A, to finally obtain k distances and choose the maximum one; which will be denoted as the Hausdorff distance [5]

For its implementation on images, the basic idea is the comparison of n vectors which have elements derived of the coordinates (x, y) that represent trajectories $S_1, S_2, ..., S_n$ from different frames, such that:

$$S_1 \Rightarrow v_1 = [a_1, b_1; a_2, b_2; ...; a_k, b_k];$$

$$S_2 \Rightarrow v_2 = [c_1, d_1; c_2, d_2; ...; c_m, d_m];$$

$$\vdots$$

$$S_n \Rightarrow v_n = [e_1, f_1; e_2, f_2; ...; e_p, f_p];$$

Notice that every vector vary on its length, due to the representation of different $M \in P(M)$. So the comparison is only valid because of Hausdorff nature. Given a trajectory S_1 as reference and once similar trajectories are found, with the aid of augmented reality the results will be displayed.

3.- Augmented Reality

Over the years, human being have been devising new ways to help describe a more intuitive and efficient way what happens in the world, everyday new technologies are created or that already have to be improved, that is the case of Augmented Reality. The Augmented Reality is a technology that in recent years has found great applications and use that goes growing up, however, the first Augmented Reality prototypes was created by computer graphics pioneer Ivan Sutherland and his students at Harvard University and University of Utha, appeared in the 1960s an used a see-through to present 3D graphics [6].

But it is up to the 90s when just beginning to return, this was due to a technology that requires considerable resources, such is the case of visualizer device, PC, besides the virtual model that would present in the real world. But, What is Augmented Reality?, it can be defined as technology that improve the visual perception of user through the overlapping of virtual objects, computer generated, into the real world. In other words, Augmented Reality allows the user see directly the real world but adding virtual objects, in this way, Augmented Reality complements the reality instead of replace completely. The figure 1 shows a line that represent where is located the Augmented Reality.

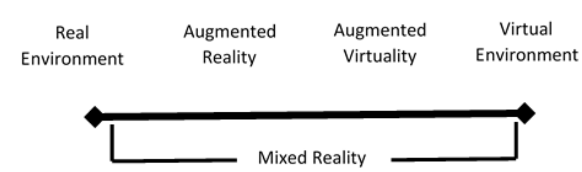


Figure 1. *Reality-virtuality continuum [7].*

An Augmented Reality system [11] [12]:

DOI: 10.2352/ISSN.2470-1173.2016.4.ERVR-422

- Combines real and virtual objects in a real environment.
- Registers (aligns) real and virtual objects with each other.
- Runs interactively, in three dimensions, and in real time.

To accomplish the above requirements a number of elements and devices needed [9]:

- Display. Is the basic element where shows the mix of virtual and real objects. Nowadays exist different elements of display (figura 2).

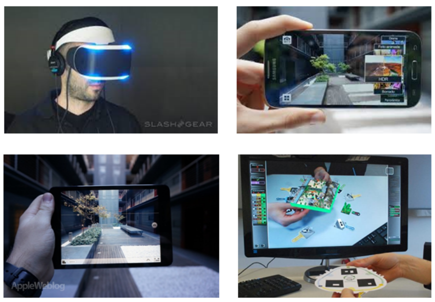


Figure 2. *Displayed devices of Augmented Reality*

- Camera. It is the device that takes the real-world information and transmits it to Display, One advantage is that no currently requires a camera with specific requirements , an example is shown in Figure 3.

Figure 3. *FOSCAM Camera*

- Software. Two kind of software are required, the fist one is used to create the virtual objects that will be displayed to real-world to create the Augmented Reality, and the second one, takes the information to mix the real-world information acquired by the camera and the created information (virtual objects).
- Marker. Normally active brands are used, however, it can works with passive brands. The brands are associated with the virtual objects that will be displayed in the real-world, helping as a positioning reference.

4.- Experimentation

To develop the experiment , a recording is performed on a stage where it had a high turnover rate ,which is ideal to find different paths and analyse this video with different approaches to query from a reference.

4.1 Motion detection

For any of the two approaches a major step is distinguish the movement regions, for this purpose temporal templates approach is used as described in [13] and [14].

$$T_{(t)}(I_t, I_{t-1}) = \alpha T_{t-1} + (1-\alpha) D_B(I_t, I_{t-1}) \qquad (3)$$

Temporal templates represents an economic way to detect motion and may be tolerant to certain luminance changes, besides, the decay function improve tolerance. For each pair of consecutive samples: T is the time decay function; α the decay constant with a $[0,1]$ domain; I_t is an image captured at time t and D_B is the movement segmentation at time t defined as:

$$D_B = \begin{cases} 1 & \text{if } d_k(I_t, I_{t-1}) \geq r\sigma \\ 0 & \text{if } d_k(I_t, I_{t-1}) < r\sigma \end{cases} \qquad (4)$$

Where $r\sigma$ is a confidence interval required to distinguish movement and is based on the probability density function (pdf) of T_t. Assuming common camera and traffic circumstances $r\sigma$ is set up to 90% of probability, as described in [5]. At last, d_k is defined as:

$$d_k(I_t, I_{(t-1)}) = abs\left[\left\|\nabla^k I_t\right\| - \left\|\nabla^k I_{t-1}\right\|\right] \qquad (5)$$

Where $k = 1$, the first order derivative difference. Further, due to the exponential properties of (3), a λ threshold is performed ($TB_t = T_t > \lambda$). As result, TB_t is the motion segmented for time t.

4.2 Temporary differences approach

This motion segmentation serves two purposes a symbolic and a template approach. The movement of an i object (M_i) is seen as a sequence of TB templates over time, $M_i = \{TB_t, TB_{t+1}, ..., TB_{t+n}\}$. So, for each M_i two paths are taken, position vectors and symbolic ones.

The first one is based on the objects centroid from motion templates at t time. This feature is useful to build position vectors for each object in scene during experimentation, based on a proximity hypothesis which states: the image displacement for an object between samples is minor to a p value.

For the second one, a symbolic approach from [13] to detect abnormal activities is used. In figure XX, the symbolic motion states(SyM) are shown for the scenario. From XX amount of experimental video several trajectories are obtained as symbolic sequences.

The approach is as follows: for each TB_t a function Ψ is applied, which give us a transformation between the (x,y) image coordinates movement representation and Σ^* via SyM, where Σ is basically the set of labels from SyM.

The coding function for that matter, is $\Phi : TB \rightarrow \Sigma^*$ and is defined as follows: $\Phi(TB) = \{SyM(x), \forall x \in TB | TB(x) = 1\}$.

Now each object template of movement can be see as a subset of SyM, hence as subset of Σ^*. The main idea is to represent M_i as a sequence of symbols s where $s \in \Sigma$. So, another function is needed, let it be $\Psi : \Sigma^* \rightarrow \Sigma$, and is defined as $\Psi(\Phi(TB))$ =mode($\Phi(TB)$). Thus, $M_i = \{s_t, s_{t+1}, ..., s_{t+n}\}$.

Nevertheless for comparison via the Hausdorff metric a function $\Phi : \Sigma \rightarrow \mathbb{N}_2$ is required. Since each motion state is a set of elements $x \in \mathbb{N}_2$ with a unique correspondence to a element $s \in \Sigma$. The position vector for each object in motion reduces its cardinality because several x positions are represented with the same symbol while the object move along the same motion state.

DOI: 10.2352/ISSN.2470-1173.2016.4.ERVR-422

However, since each motion state is associated with only one symbol on a bijective fashion and each $r_i = \{x_1, x_2, ..., x_n | x \in \mathbb{N}_2\}$. $\Phi : \mathbb{N}_2 \times \mathbb{N}_2 \times ... \times \mathbb{N}_2 \rightarrow \mathbb{N}_2$ instead. $\Psi(r_i) = max\{d(SyM_{r_i})\}$, where d is the morphological distance transform and SyM_{r_i} is the binarized r_i region of SyM as $SyM(x)_{r_i} = 1, \forall x \in r_i$ and $0\ \forall x \notin r_i$.

So when applied Φ, it will return something less expressive than the original position vector, due to the general motion pattern detection approach for abnormal activity detection establish on [14].

5.- Results

Similar movement dynamics, as result of different approaches, deliver distinct information creating heterogeneous data bases. Leading to this proposal, where a new way to compare that information is introduced. The comparison is realized through the characteristics extraction, which represents different trajectories obtained by different approaches. Once the characteristics are extracted, a comparison is performed using the hausdorff distance to find similar trajectories by a visual query-by-example.

Based on the approaches described in the previous sections, in order to analyse the trajectories obtained Tr_{s1}, Tr_{s2}, within which the similar trajectories are obtained to a visual query-by-example Tr_{ref} defined from a manually drawn trajectory. Figure 4 represents a video frame to exemplify the scenario under experimentation took place, the visual query-by-example Tr_{ref} is presented in figure 5. For each used model $(M \in P(M))$, the centroid is calculated to obtain vectors of different length n found for each trajectory of each model. The results of the fist approach S_1 is presented in figure 6, finally the result of the second approach S_2 is presented in figure 7.

Figure 4. *Experimental Scenario*

Note that the trajectories obtained by the two approaches are different because the frame rate are different for each one.

To obtain the similar trajectories firstly applies the Hausdorff distance H_d, once the result is obtained similarity is determined by applying a threshold U_1. If $H_d < U_1$ then $Tr_m \in S_n \sim Tr_{ref}$.

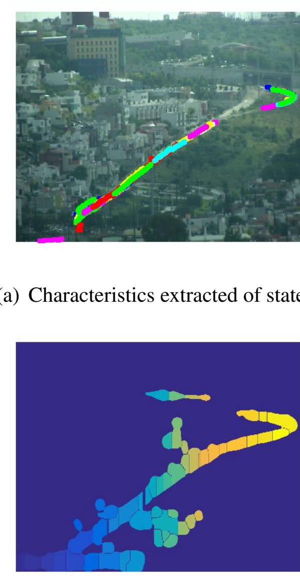
If a very high threshold is defined the similar trajectories increases, if a very low threshold is defined, the number of similar trajectories decreases. The threshold determined to find the closer trajectories is represented by $U_1 = 80$, the figure 8 presents the obtained result. In each obtained result the query-by-example approach is represented by the yellow color.

To complement the experimentation, it has defined a new threshold U_2, with which it seeks to demonstrate that the increase it increases the amount of similar trajectories, figure 9, presents the obtained result. The threshold used is defined by $U_2 = 100$.

Finally, a third threshold is defined $U_3 = 120$, the amount of similar trajectories increases, the obtained results is presented in figure10.

The same values for the threshold $U_1, U, 2, U_3$ defined previously are used to get the similar trajectories for the second ap-

Figure 5. *Query-by-example approach*



(a) Characteristics extracted of states

(b) Total States.

Figure 6. *Motion detection.*

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.4.ERVR-422

proach S_2, the obtained results are shown in figures 11, 12, 13.

Table 1 presents the used approach, the threshold defined and amount of similar trajectories obtained.

Results

S_n	U_x	Tr_{sim}
S_1	80	5
S_1	100	8
S_1	120	19
S_2	80	2
S_2	100	9
S_2	120	14

Finally, obtained results are visualized aided by Augmented Reality to display the similar trajectories obtained applying the Hausdorff distance, from a reference with a threshold defined.

(a) Characteristics extracted of Temporary Differences approach.

(b) Temporary Differences approach.

Figure 7. *Temporary Differences approach.*

Figure 8. *Similar trajectories* (S_1) $U_1 = 80$.

Figure 9. *Similar trajectories* (S_1) $U_2 = 100$.

Figure 10. *Similar trajectories* (S_1) $U_2 = 120$.

Figure 11. *Similar trajectories* (S_2) $U_2 = 80$.

DOI: 10.2352/ISSN.2470-1173.2016.4.ERVR-422

6.- Conclusions

By having a variety of ways to define dynamics of movement on stage, it creates a lot of information that is heterogeneous with each other, to make comparisons between these different approaches it is necessary define a criterion that be useful to find similarities and display results aided by Augmented Reality

The goal of using Augmented Reality to presents results is to show the trajectories that are similar and extract more information from the scene to get more complete conclusions combining the results of different approaches used.

This work presents a new way to realize comparisons of heterogeneous information, proving that using the hausdorff distance similarities can found on trajectories, extracting characteristics points that represents each trajectory obtained by different approach, efficiently.

Acknowledgements

We would like to thank our colleagues, $CIDESI_{TM}$ and ASA-CONACYT project with number ASA-2014-1-242864 for their support, with regarding the infrastructure for the development of this work.

Figure 12. *Similar trajectories (S_2) $U_2 = 100$.*

Figure 13. *Similar trajectories (S_2) $U_2 = 120$.*

References

[1] Juan Manuel García Huerta, Generación Automática de Modelos Gramaticales para la inferencia de actividades en un Sistema de Visión, Master Thesis. (2014).

[2] Shian-Ru Ke, Le Uyen Thuc Hoang, Yong-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo, and Kyoung-Ho Choi, A review on video-based human activity recognition, Computers, 2(2):88, 131. (2013).

[3] C. Kosniowski, Topología Algebraica, @Editorial Reverté, S. A., 1986, pg. 55.

[4] Zezula, P., Amato, G., Dohnal, V., & Batko, M.. Similarity search: The Metric Space Approach (Vol. 32). Springer Science & Business Media, 2006, pg. 220.

[5] Fuentes Henry Arguello, Comparación de Huellas Dactilares usando la Distancia de Hausdorff, Sistemas, Cibernética e Informática, 5, 35 (2008).

[6] Tamura Hideyuki, Steady steps and giant leap toward practical mixed reality systems and applications, Proc. Int. Status Conf. on Virtual and Augmented Reality, Leipzig. pg. 3. (2002).

[7] Mistry P., Maes P., Chang L. WUW-wear Ur world: a wearable gestural interface. In CHI'09 extended abstracts on Human factors in computing systems. pg. 4111. (2009).

[8] Van Krevelen, D. W. F., R. Poelman. A survey of augmented reality technologies, applications and limitations. International Journal of Virtual Reality, 9, 1 (2010).

[9] Carlos Machado Marcos, Aplicación de la Realidad Aumentada para un sistema de entrenamiento. (2011).

[10] Mubarak Shah, Ramesh Jain, Motion-Based Recognition, Springer, 1997.

[11] R. T. Azuma. A survey of augmented reality. Presence, pg. 355, (1997).

[12] R. T. Azuma, Y. Baillot, R. Behringer, S. K. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. IEEE Computer Graphics and Applications, 21,34 (2001).

[13] J.M. García-Huerta, Hugo Jiménez-Hernández, Ana-M Herrera Navarro, Teresa Hernández-Díaz and Ivan Terol-Villalobos, Modelling dynamics with context-free grammars, IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, pg. 902611 (2014).

[14] Hugo Jiménez-Hernández, Jose-Joel González-Barbosa, and Teresa Garcia-Ramírez. Detecting abnormal vehicular dynamics at intersections based on an unsupervised learning approach and a stochastic model. Sensors, pg. 7576 (2010).

Biography

Josué Rafael Montes received his BSc. in Mechatronics Engineering from the Universidad del Valle de México (UVM), México, 2009. Aspiring to a PhD. degree in Mechatronics at CIDESI, in Querétaro, México. His work has focused on Automatic Computational Modelling of Human Behaviour and Visual Surveillance through Multi-camera Vision System aided by Augmented Reality.

Hugo Jiménez Hernandez, PhD in Computer Vision. in CICATA-IPN, México, MS in Computer Sciences in CIC-IPN, México, Bachelor in Computer Sciences in ITQ-México. His current interest includes Automatic Learning, Computer Science Theory and Automatic Visual Inferences. Nowadays He is the head of computer Vision Lab at CIDESI-México. He has around of 50 works in several international journals and congress.

DOI: 10.2352/ISSN.2470-1173.2016.4.ERVR-422

J. A. Soto-Cajiga (M'11) received the BEng. degree in Control and Instrumentation Engineering from the Instituto Tecnológico de Querétaro, México, in 2003, the MEng. degree in Mechatronics from the PICYT (Inter-institutional Graduate in Science and Technology), México, in 2006 and the PhD. degree also from PICYT in 2012. He works for CIDESI (Engineer Center and Industrial Development), as researcher in the applied research area, designing embedded system and real time signal processing algorithms. He is also professor of graduate level in PICYT, and his research interests include real-time signal processing implemented in hardware, algorithms parallelizing for signals and video processing and electronics for ultrasonic PIGs.

Leonardo Barriga Rodríguez, PhD in Mechatronics in CIDESI, México; with 10 years of experience in software development, computer vision, image processing and inspection robotics. He has participated in the development of more than 20 research projects in the industry. He is currently a research professor of CIDESI (Engineer Center and Industrial Development)).

Teresa Hernández Díaz received the BSc. in Mechatronics Engineering at Universidad Politécnica de Sinaloa (UpSin), México, 2009. Currently a candidate to a PhD. degree in Mechatronics at CIDESI, in Querétaro, México. Nowadays, her recent researches focus on automatic environment modelling through a laser array system, detecting dynamic object's interaction and associative memories.

Juan M. García Huerta received the BSc. in Mechatronics Engineering at Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM), México, 2011. Aspiring to a PhD. degree in Mechatronics at CIDESI, in Querétaro, México. His research interests include automatic computational modelling of humans behaviour and visual surveillance.

Alberto Vázquez Cervantes received the BSc. in Electronic Engineering at Instituto Tecnológico de Querétaro, México, the MEng. degree in Mechatronics from PICYT (Inter-institutional Graduate in Science and Technology), México. Currently a candidate to PhD. degree in Mechatronics at CIDESI, in Querétaro, México. His research interests include artificial intelligence combining with embedded systems.