Active tuning of intrinsic camera parameters

Christian Micheloni, *Member, IEEE*, Gian Luca Foresti, *Senior Member, IEEE*

***Abstract*—In the last years, the research effort of the scientific community to study systems for ambient intelligence has been really strong. Usually, the systems developed so far base their analysis on images acquired by automatic cameras. In this paper, we propose a way to develop new smart systems that are able to actively decide both what to see and how to see it. In particular, the main idea is to tune the acquisition parameters on the basis of what the system desires to acquire. The regulation strategy is based on two camera parameters, focus and iris. It aims to identify an optimal sequence of steps to enhance the acquisition quality of an object of interest. To this end, a hierarchy of neural networks has been employed first to select which parameter must be regulated then to adjust it. The proposed solution can be applied to both static and moving cameras. The results show how the proposed technique can be applied to images acquired by a moving camera with zoom capabilities for surveillance purposes.**

I. Introduction

Modern visual-based surveillance systems [1]–[4] base their analysis on the detection of moving objects. Different application domains require different object detection techniques and system settings. In the transportation field, vehicles are the main object of interest for traffic monitoring purposes [5], [6]. Persons are fundamental for human activity analysis [7], [8], then trajectories computed on such objects can be exploited for behaviour understanding [9]. Usually, video surveillance applications imply to pay attention to wide areas, hence different kinds of cameras are generally used, e.g, fixed cameras [10], omnidirectional cameras [11] or moving cameras [12]–[16]. Recently, systems using moving cameras have been widely considered thanks to their capacity of patrolling large sections of the monitored area. Moreover, systems using these cameras can look at what they need simply by controlling the motion parameters to redirect the camera's gaze. In the literature, such systems are considered as belonging to the field of *Active Vision* [17].

For all image processing systems, the quality of the acquired image is an important feature for the image formation process. In this research, we focus our attention on the determination of the intrinsic camera parameters to enhance the image acquisition quality. The idea behind this work is that controlling the image acquisition quality is preferable than simply assessing it [18]. Managing the image formation process with respect to the current image processing tasks is better than designing complex and robust image processing modules operating on low quality images. This exactly fits with the definition of *Active Vision*. More, such a concept is extended in such a way that *Active Vision* is not just the interaction among the observer and the sensors to actively decide what to see [17] but also *how* to see it.

Following this new concept, *Active Vision* should mean more than just using moving sensors. We want to expand this concept by including techniques that allow to *actively* control the quality of the image acquisition with respect to an object of interest for image processing purposes. Actual digital cameras consider only a restricted area around the centre of the image to tune the optimal intrinsic parameters. In addition, such techniques have the objective to improve the quality for the humans perception. With this research we want to change this perspective by keeping the objects of interest at the centre of the tuning process. Hence, the proposed solution could be useful in all the ambient intelligent tasks in which decisions have to be taken automatically by intelligent processes. These will take their decisions not on the basis of the perception quality but on the objects' appearance quality.

To achieve such a result, the proposed method adaptively tunes two acquisition parameters (i.e., focus and iris) by applying quality operators on the object of interest. The control strategy is based on a hierarchy of neural networks trained on quality operators values and camera parameters.

II. Related Work

To better clarify the purposes of this research it is important to understand which are the characteristics of good quality images. Typically, good quality images have a high sharpness degree with respect to focused objects [19]. If we analyse these objects in the frequency domain, we see how their edges are described by high frequency components while smooth (defocused) zones are described by low frequencies [20], [21]. Therefore, since images with high sharpness degree mean focused images, the modern cameras technology aims to maximise the high frequency components [22], [23].

Krotkov in [24] presented and tested various criteria for computing the defocusing degree. Such criteria are all based on quality functions that effectively play an important role in determining the complexity and the efficiency of the solution. These functions should present a set of desired characteristics to be efficiently employed in a regulation strategy:

- The function should achieve the maximum (minimum) value when the best quality is reached.
- The function should not present local maxima (minima).
- The function should be independent of the structure of the objects inside the image.

In addition, the algorithms for increasing the image quality should be evaluated considering the following rules:

- Unimodality: the function should have just a maximum (minimum).
- Amplitude difference: the quality function should present significant variation between optimum and poor quality.
- Range amplitude: the quality function should continuously vary inside a considerable range of values.

- Complexity: the implementation should be as fast as possible.
- Generalisation: the solution should not consider or be limited to particular cases.

The regulation process here proposed, to guarantee the desired properties, is based on well known quality operators like *tenengrad, flatness, entropy, saturation, max/min grey levels, luminance, maximum local difference, variance, etc.* [25], [26]. In particular, the work introduced by Murino *et. al* [27], which uses different quality operators [28], has been taken into account for the proposed solution. The operators are linearly combined to produce a classification function for the assessment of the image quality. The quality degree of an image is given by a global quality monotonic descendant function Q able to carry information about the global quality of the image. To enhance the image quality, Q has to be minimized.

The proposed strategy iteratively tunes an intrinsic parameter among Focus, Iris aperture, Gain and Black level. Once a parameter is set to a new position, a new image is acquired and a new Q value is computed. If Q has decreased, the change is accepted. Otherwise, the change is discharged and a new parameter is chosen. This problem can be seen as the search of the minimum value of a function whose domain is represented by the hyperplane defined on the camera parameters ranges.

The impossibility to estimate the surface of the function Q avoids the use of gradient descendant algorithms. Murino *et al.*, by analysing the situation that may occur during the acquisition process, identified that the image quality can be generally classified into two main categories: a) out of focus and b) with a bad brightness. Moreover, a relation between the focus or brightness quality and the four intrinsic parameters has been derived.

This yielded to define a strategy based on extracting simple features in order to choose the parameters to set. The general strategy is based on the definition of two *semi quality* functions defined as linear combinations of simple operators. A function called *Brightness Quality* (BQ) is introduced to estimate the brightness degree and a function called *Sharpness Quality* (SQ) is introduced to estimate the focus degree (see [27] for the definitions of BQ and SQ).

Thus, to complete the regulation strategy Murino *et al.* adopted different thresholds TH_S, TH_{B1} and TH_{B2} to establish if an image is well focused and/or has a correct brightness. In general, an image has a good focus degree if $SQ > TH_S$, while if $TH_{B1} < BQ < TH_{B2}$ then the image has a correct brightness. Depending on the values of the two functions three situations may occur:

1) an image is out of focus and its brightness is correct $\Rightarrow$ activation of the focusing strategy.
2) an image has a good focus degree and its brightness is wrong $\Rightarrow$ activation of the brightness strategy.
3) an image is out of focus and its brightness is wrong $\Rightarrow$ activation of the brightness quality

While the control of the brightness could be deterministic (lighter or darker), the authors state that there is no information for deciding how to adjust the focus parameter (e.g. near or far). Summarising, the method proposed by Murino *et al.* presents two main limits:

- The determination of the optimal values for the thresholds TH_{B1}, TH_{B2}, TH_S
- The development of an efficient strategy for the regulation of the parameters.

The use of fixed thresholds strongly depends on experimental tuning, on the conditions of the experiments, on the context of the acquisition, on the adopted sensor, etc. In addition, the use of a random strategy to decide the focus direction (far or close), even though supported by a trial-and-error technique, does not represent an optimal solution. A deterministic solution for the automatic focus parameter regulation is still missing. As matter of fact the number of steps to determine the optimum focus is a key issue. In [29], Kehtarnavaz and Oh present a new rule based approach that speeds up the search for the focus peek. With respect to the trivial global search and to a binary search the proposed solution achieves considerable improvements. In particular, the number of steps to compute the optimal focus is reduced to be around 61 on the average.

In the remainder of the paper, a description of the proposed system is presented in Section III. In Section IV, the strategy to decide how to tune the intrinsic parameters is presented. Finally, in Section V a deep validation of the proposed solution showing how the introduced novelties improve the actual solutions is given.

III. System Description

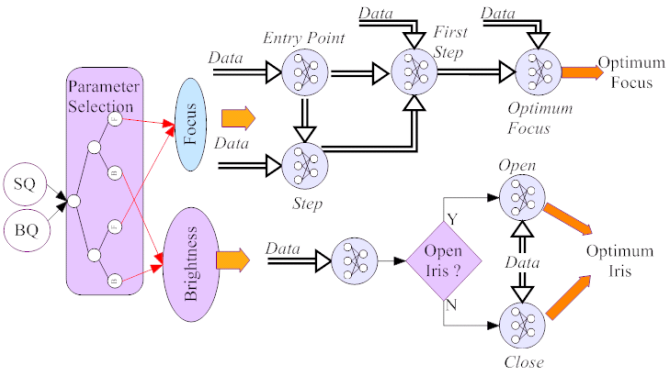


Fig. 1. General architecture of the regulation strategy. The first step identifies the parameter to fix, thereafter the proper set of networks is activated to compute the correct value of the selected parameter.

A CCD progressive colour camera mounted on a Pan-Tilt platform is used to acquire images of the monitored area. Within the controlled environment, there are no restrictions either in terms of number of moving objects or in terms of their movements. The object detection is assigned to a low level change detection module adopting the registration technique proposed by Micheloni and Foresti in [30]. A history of the moving objects is stored in the states of a finite automaton which allows to maintain track of all the objects of interest inside the environment. Within the set of objects, a high level module identifies an object of particular interest. Its blob is then considered as the area of the image on which the proposed regulation strategy is applied.

The objective is to overcome the limits in [27] by proposing two major improvements on the strategy regulation:

- Exploitation of a neural tree to determine if an image is out of focus and/or it presents a non optimal brightness.
- Development of a hierarchy of neural networks for a fast computation of the optimal focus and iris positions for focusing and tuning the brightness.

In particular, as shown in Fig. 1, first a Generalised Neural Tree (GNT) [31] is used to decide whether to tune the focus or to improve the brightness. Thus, such a tree operating as a classifier, identifies whether the image area needs to be focused or needs a different brightness. Comparing this novelty with the method proposed in [27], there is no need to adopt any ad hoc fixed threshold for biasing the tuning strategy. In addition, the generalisation of the tree is guaranteed by the heterogeneity of the samples introduced in the training set. This allows to use the developed solution in different contexts without requiring any further tuning of the thresholds.

The considered GNT has been trained by patterns composed by couples of BQ and SQ values. The possible classifications are represented by the set C={out-of-focus, wrong brightness, out-of-focus with wrong brightness}.

Once the classification has been performed, two different branches can be taken for the regulation of the considered intrinsic parameters. If the GNT demands a focus regulation, four neural networks, based on the tenengrad measure, are used. The definition of the tenengrad operator is based on:

$$|\nabla I(x,y)| = S(x,y) = \sqrt{I_x^2 + I_y^2} \qquad (1)$$

where I_x and I_y are the gradients of the image respectively on the x and y directions. The tenengrad operator is then given by

$$TN = \sum_{x=1}^{N}\sum_{y=1}^{N} S(x,y)^2 \qquad for \;\; S(x,y) > Th \qquad (2)$$

where N is the number of pixels belonging to the area of interest and Th is a threshold. In the current work, such a threshold is automatically computed by using the thresholding technique introduced by Kapur [32].

The objective is to identify four points within the focus range in order to estimate an optimal focus position. Thus, the required four focus regulations guarantee a considerable improvement if compared to the 61 steps needed by the technique proposed in [29].

On the other hand, if the GNT demands a brightness regulation, a first neural network is used to identify whether the iris must be opened or closed. Then, specific networks for each of the two cases have been trained to identify the optimal iris position for the next time instant.

With respect to commercial solutions, the proposed solution allows to restrict the area of interest as desired (e.g. the bounding box of a selected object). This feature applied on a target tracked by an active method [30] allows to maintain the gaze on an object of interest by keeping an optimal quality of its acquisition. This is a real breakthrough. We are proposing a new *Active Vision* paradigm in which the observer (i.e. the system) totally tunes the intrinsic (focus and iris) and extrinsic (pan, tilt and zoom) camera parameters to optimally acquire what the system requires.

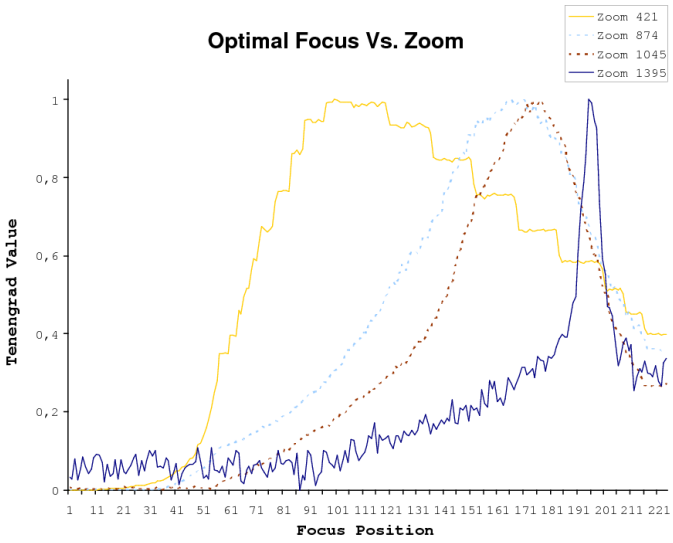


Fig. 2. The chart plots the tenengrad computed on an object's blob. The values have been obtained by sliding the focus position through its entire range for four different zoom levels. The environmental conditions, the object and the camera position have been kept constant.

In the following section, the attention will be focused on the techniques to tune the selected intrinsic parameters.

IV. Focus and Iris Tuning

The tuning of the parameters takes into account two intrinsic camera parameters: the focus and the iris. To tune each of these two parameters, two different neural network hierarchies have been studied. Since the tuning process of one value does not locally (for each tuning step) depend on the other parameter, the training processes of the two hierarchies have proceeded separately. In the following subsections, the two processes are presented.

A. Focus Regulation

As shown in Fig. 1, the adopted strategy employs four different neural networks (NNs) that we called respectively: *Entry Point, Step, First Step, Optimum Focus*.

Studying the tenengrad operator with respect to different environmental contexts and camera configurations, it is noticeable how such an operator is very noisy. In addition, since it is basically a sharpness measure, it does not allow to infer information about the optimal focal position. In particular, from a single measure, it is not possible to determine the direction (far/near) for the regulation. Therefore, it is mandatory a sequence of measurements in order to determine the right direction.

In Fig. 2, different tenengrad functions with respect to different focus and zoom positions are shown. It is clear that the optimal focus position of the same object varies when the zoom changes. This is due to the fact that the depth of field is linked to the focal length (zoom) of the optics. What is interesting to notice is how the *bell shape* of the tenengrad function narrows when the zoom level increases

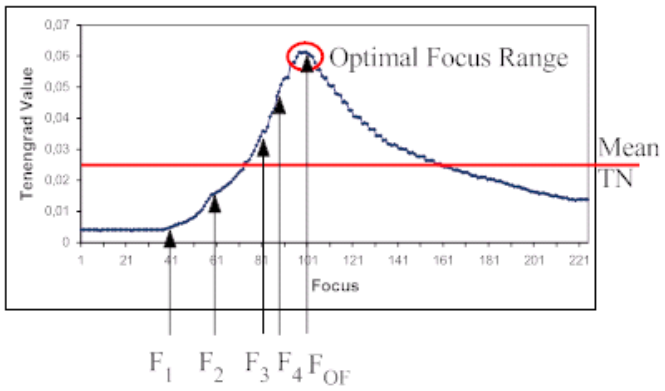


Fig. 3. To compute the optimal focus two focus values (e.g. TN_3 and TN_4 at position F_3 and F_4 respectively) greater than the threshold TN are needed. If the current tenengrad value is smaller than the threshold, two points (e.g. TN_1 and TN_2 at position F_1 and F_2 respectively) falling below such a threshold are required to determine the F_3 and F_4 points. The threshold TN has been computed as the mean of all the tenengrad values determined for all the training sequences. In such a way, such a threshold can be considered as the mean value of the tenengrad operator.

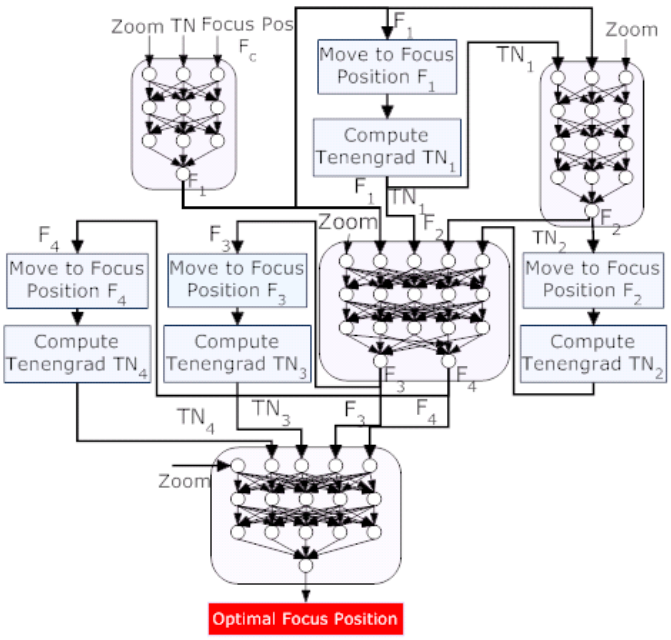


Fig. 4. Network hierarchy and strategy for the tuning of the focus parameter.

(lowest zoom 421 vs. highest zoom 1395). This means that, as we acquire an object with a higher zoom level, we must pay more attention to the focus positions. In these cases, a small displacement from the optimal position results in a big variation of the sharpness quality. Another parameter that influences the depth of filed is the iris aperture. While a zoom operation changes the optimal focus position the iris aperture does not. This analysis suggested to introduce the zoom position as a valuable parameter for the tuning of the focus. Instead, the iris position has not been considered. Anyhow, both zoom and iris parameters cannot be tuned while a focusing regulation is active.

In Fig. 3, the meaning of the proposed strategy is shown. The tenengrad measure computed on an object of interest is plotted. The horizontal line, called *Mean TN*, represents the mean value for the tenengrad measures computed on different objects acquired with the same zoom level. Thus, for our purposes, such a line represents a threshold for deciding if the current tenengrad measure describes an object really out of focus or an object that needs a fine focus tuning.

For the first case, we have developed the following four steps strategy:

- Compute, inside the entire range, a first focus position F_1 that falls below the threshold TN.
- Given the entry point F_1, move the focus into a new position F_2.
- Based on the slope of the line passing through F_1 and F_2, compute two new focus positions F_3 and F_4 that reside above the threshold TN.
- Based on the slope of the line passing through F_3 and F_4, compute the optimal focus position F_{OF}.

In the second case, when the object is not really out-of-focus (i.e. the current tenengrad measure is greater than the threshold TN), we just need to get two points above the threshold TN before running the last step of the aforesaid strategy. Since the current measure is already above TN, a small motion of the focus is performed for getting the second point. In particular, from the current position the focus is moved nearer into F_n position and farther into F_f position. The second point F_4 is therefore chosen as follows:

$$F_4 = \begin{cases} F_f & \text{if } TN_f \geq TN_n \\ F_n & \text{Otherwise} \end{cases} \quad (3)$$

where TN_x is the tenengrad value computed at focus position F_x.

To develop such a tuning strategy, a neural network hierarchy has been developed (see Fig. 4). If the object is really out-of-focus (i.e. the tenengrad value is lower than a threshold), the *entry point* NN is applied on the current focus position F_c, its related tenengrad measure TN_c and the current zoom value. The developed NN, defined by means of a trial and error procedure, is full connected and composed by three input nodes, two hidden layers each composed by three nodes and an output node. The output value determines the first focus position F_1 within the focus range.

At this point, the system moves the lens to reach the focus position F_1. Once the repositioning is achieved, a new tenengrad measure TN_1 is computed over the area of interest. This new tenengrad value, the corresponding focus position F_1 and the zoom value are given as input to a second NN called *Step*. This is also full connected and composed by three input nodes, three hidden layers with three nodes each and an output node. Such a NN finds out a second focus position F_2 whose tenengrad value falls below the threshold TN. The camera is reconfigured to reach such a focus position for which a new tenengrad measure TN_2 is computed.

After these two first steps, it is possible to determine the slope of the line connecting the two focus positions in order to estimate two new values greater than TN. For such a purpose, a further NN takes as input the pattern $(F_1, TN_1, F_2, TN_2, Zoom)$ given by the two computed focus

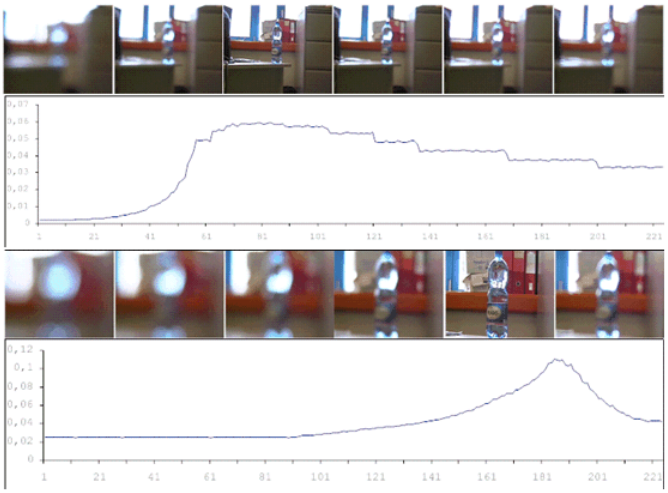


Fig. 5. Examples of indoor sequences belonging to the training set. For each sequence the corresponding tenengrad chart is plotted. In this case, the object of interest is represented by a bottle of water.

positions with the corresponding tenengrad measures and the zoom position. This NN, called *first step*, is full connected and composed by five input nodes, two hidden layers with five nodes each and two output nodes. These, determine the two focus positions F_3 and F_4.

Computed the tenengrad measure for these two last focus positions, the strategy reaches its last step. The last *optimum focus* NN is executed on the pattern $(F_3, TN_3, F_4, TN_4, Zoom)$ given by the two focus positions, the corresponding tenengrad measures and the zoom position. Such a NN is full connected and composed by five input nodes, two hidden layers with five nodes each and an output node. The value returned by the output node is the *optimal focus* position F_{OF} that improves the sharpness value computed on the object of interest.

If at the next step a new focus regulation is requested, the last two focus positions F_4 and F_{OF} are chosen as F_3 and F_4. It is interesting to notice how, with this strategy, just four focus regulations are required to reach the optimal value.

Training Focus Network: Once the hierarchy of the four NNs has been defined, the training phase has followed. At this point, the definition of the training set is crucial to achieve good performances. In particular, to generalise the solution to different illumination conditions, both indoor and outdoor training sequences have been acquired. Each recorded sequence consists in a frame for each possible focus position within the focus range. Precisely, 40 (10 scenes, 4 zoom levels) indoor sequences and 80 (20 scenes, 4 zoom levels) outdoor sequences have been used. In Fig. 5 some frames of the training sequences obtained by acquiring the same scene using different zoom levels are shown.

Once the sequences for the training set definition have been acquired, the tenengrad values for the all frames of the training sequences have been computed. The mean value ($MeanTN$) (i.e., the mean tenengrad value of the training sequences) represents a threshold between good and pour focus positions. When $MeanTN$ is plotted together with a tenengrad chart (see Fig. 3), the pour focus positions generate two tails. One on the left side of the chart and one on the right. For the clarity of the presentation, hereafter only the left tail of the chart is considered. Though, the same applies to the right tail.

To train the *Entry Point* neural network, for each sequence $i = 1, \ldots, 120$, the tail of the tenengrad chart, given by the range that goes from the nearest focus position to the point F^i_{TN}, has been considered (see Fig. 6(a)). The tenengrad value computed on F^i_{TN} is the greatest value smaller than the $MeanTN$ value. Within such a range, a focus position F^i_{rnd} has been randonmly selected and its tenengrad value TN^i_{rnd} has been computed. These two values and the zoom position $Zoom^i$ have been included in the training set as a input pattern. Therefore, the mid point F^i_1 of the range $[nearest, F^i_{TN}]$ has been selected as the desired output focus position. The selection of the mid point follows a binary search strategy also adopted in [29]. Summarising, the training set for the first *Entry Point* NN is defined as follows:

$$TS_{EP} = \{((F^i_{rnd}, TN^i_{rnd}, Zoom^i), F^i_1) | i \in [1..120]\} \quad (4)$$

It is worth noticing that the training set is composed by data related to focus positions whose tenengrad values are smaller then $MeanTN$. The same holds for the desired output values. The stop criteria adopted for the training process has consisted in a maximum expected error of 0.01 equivalent to $1/100$ of the minimum focus step. Such a stop criteria required 9×10^4 training epochs to converge to a solution.

For the training of the *Step* NN, the operative situation has been considered (see Fig. 6(b)). In particular, the *Step* NN is executed on a pattern that is determined by the *Entry Point* NN (i.e. F_1). Thus, to train the *Step*, for each sequence i, a focus position $F^i_{rnd} \in [nearest, F^i_{TN}]$ has been randomly selected. Such a value defines the input pattern for the *Entry Point* NN whose output is the focus position $F^i_{1^{EP}}$. For each sequence i, the position $F^i_{1^{EP}}$ represents a good approximation of the optimal output F^i_1 associated to the previous NN. So, the input pattern of the *Step* network is determined on the basis of the $F^i_{1^{EP}}$ focus position. To define the desired output value, following the strategy adopted in the previous training process, the mid point F^i_2 between $F^i_{1^{EP}}$ and F^i_{TN} has been selected. Hence, the training set for the *Step* NN is defined as follows:

$$TS_S = \{((F^i_{1^{EP}}, TN^i_{1^{EP}}, Zoom^i), F^i_2) | i \in [1..120]\} \quad (5)$$

It is worth noticing that, even in this case, the training set is composed by patterns containing focus positions whose tenengrad values are smaller than $MeanTN$. In addition, the input patterns are composed by values computed by the first NN as it will be in an operative situation. For the training process of this NN, the maximum expected error has been set to 4×10^{-3}. Such a stop criteria required 9.6×10^4 training epochs to converge to a solution.

To train the *First Step* NN, the same concept used for the *Step* NN has been adopted (see Fig. 6(c)). To determine the input pattern, for each sequence i, a focus position F^i_{rnd} has been randomly selected and given in input to the *Entry Point* NN. Its output $F^i_{1^{EP}}$ and the related values (i.e. $TN^i_{1^{EP}}$ and

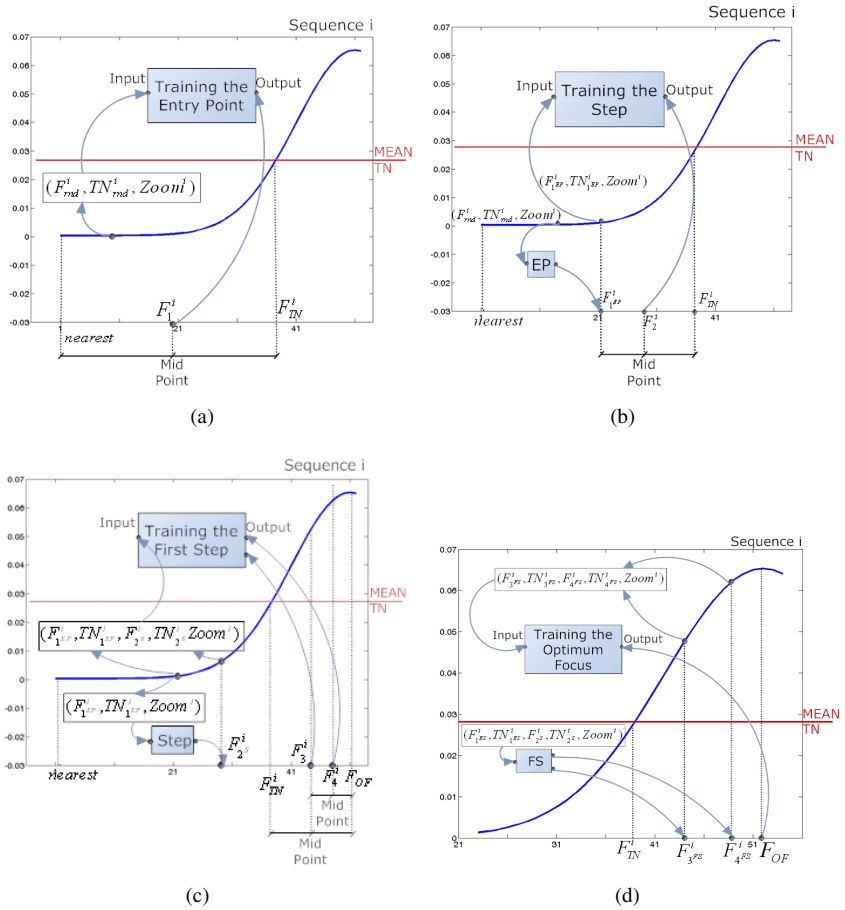


Fig. 6. Convergence plots of the four Neural Networks developed

$Zoom^i$) have been given in input to the trained *Step* NN. The output $F^i_{2^s}$ of this second NN represents a good approximation of the optimal value F^i_2 determined for the sequence i during the definition of the previous training set.

At this point F^i_{1EP} and $F^i_{2^s}$ represent two focus positions whose tenengrad values are minor than $MeanTN$. Hence, looking at the focusing strategy, they represent two good positions to train the third NN *First Step*. To determine two optimal output positions, the mid point F^i_3 of the range $[F^i_{TN}, F_{OF}]$ and the mid point F^i_4 of the range $[F^i_3, F_{OF}]$ have been considered. Thus, the range related to the left side of the good focus values has been iteratively bisected. F^i_3 and F^i_4 are focus positions whose tenengrad values are both greater than $MeanTN$. Therefore, the training set for the *First Step* NN can be defined as follows

$$TS_{FS} = \{((F^i_{1EP}, TN^i_{1EP}, F^i_{2^S}, TN^i_{2^S}, Zoom^i), F^i_3, F^i_4) | i \in [1..120]\} \quad (6)$$

Let be $\{((F^i_{1EP}, TN^i_{1EP}, F^i_{2^S}, TN^i_{2^S}, Zoom^i)\}$ a pattern determined by randomly selecting a focus position F^i_{rnd} on the sequence i and by running the *Entry Step* and *Step* NNs on it (see Fig. 6(d)). Such a pattern given in input to the *First Step* NN generates two focus position F^i_{3FS} and F^i_{4FS}. These two positions define the input pattern of the last NN. Obviously, to train the *Optimum Focus* NN, the desired output value is the position F_{OF} whose tenengrad value is the maximum. Thus, to train the *Optimum Focus* NN the training set has been defined as:

$$TS_{OF} = \{((F^i_{3FS}, TN^i_{3FS}, F^i_{4FS}, TN^i_{4FS}, Zoom^i), F^i_{OF},) | i \in [1..120]\} \quad (7)$$

For the training process of the last two NNs, the maximum expected error has been set to 2×10^{-3}. Such a stop criteria required about 10×10^4 and 11×10^4 training epochs to converge to a solution respectively for the *First Step* and *Optimum Focus* NNs.

As it can be understood, the adopted strategy for the definition of the training set is not trivial. As matter of fact it is a synthesis of several tests as much as selection methods. In particular, selecting the training sets autonomously for each network (i.e. without considering the output of the preceding NNs) yielded to the missed convergence of the NNs or to unpredictable outputs. For the convergence, the high variability of the randomly selected input patterns did not allowed to reach the required training errors. To achieve the convergence, greater expected errors have been necessary. This solution yielded to NNs whose outputs were not in accordance with the defined strategy. In particular, the *Step* network computed focus positions whose tenengrad values were greater than $MeanTN$. Similarly, the *First Step* computed values were either not greater than $MeanTN$ or on either side of the optimal point. These unpredictable values do not allow to reach a position that well approximates the optimal focus. Instead, by adopting a cascade strategy, the patterns variability

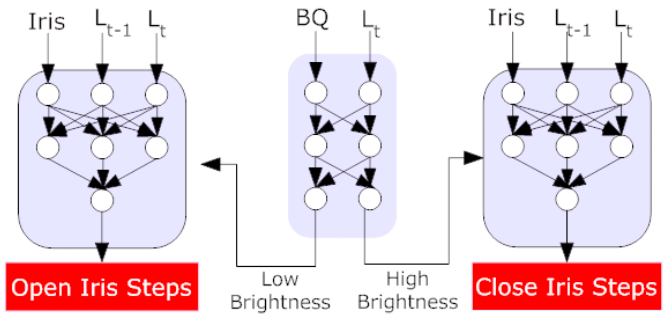


Fig. 7. Network hierarchy and strategy for the iris tuning

Fig. 8. Sample images belonging to the training sequences. The bounding boxes determine the image area considered for the computation of the parameters of interest.

is reduced and the convergence is achieved. In addition, the *Step* NN gives always focus positions whose tenengrad values are smaller than $MeanTN$. Similarly the *First Step* NN guarantees two focus positions on the same side with respect to the optimal focus. Thus, the strategy is always guaranteed to operate on expected values.

B. Iris Regulation

Compared to the focus tuning, the regulation of the brightness is much easier. Whilst the tenengrad value does not suggest the tuning direction, the brightness does. From the classification of the current object brightness as too dark or too bright, it is possible to decide if the iris should be opened or closed. Hence, the regulation strategy (see Fig. 7) for the brightness value takes into account three different feed forward NNs trained with a back propagation algorithm. The first NN is composed by two input nodes, a hidden layer with two nodes and two output nodes. This is the network that classifies the image region as too dark or too bright. The input pattern is composed by the value of the semi-quality function BQ and by the Luminance value defined as follows:

$$L = \sum_{\mathbf{x} \in B_i} V(\mathbf{x}) \tag{8}$$

where B is the blob of the considered object and $V(\mathbf{x})$ is the *Variance* value computed on the position of the pixel $\mathbf{x}$ and on the HSV colour model. The *Variance* is computed by considering the grey level as a stochastic variable. It is given by the quadratic difference between the grey values and their mean μ:

$$V = \frac{1}{N^2} \sum_{x=1}^{N} \sum_{y=1}^{N} [I(x,y) - \mu]^2 \tag{9}$$

The desired output of the network is the classification of the blob brightness as low or high.

The classification of the blob as too dark or too bright allows to run the proper NN to open or close the iris. As shown in Fig. 7, both NNs, to estimate the optimal iris position, use a input pattern defined by the current iris position and by the luminance values at the current time instant t and at previous time instant $t-1$. The difference between these two NNs is that one is trained to open the iris while the other to close it. The outputs of the NNs is the number of closing or opening iris steps.

To train the NNs the strategy followed was the same adopted to train the focusing NNs. Thus, 20 sequences have been acquired by scanning all the iris positions within the range. This means that, by acquiring 30 images for each sequence, a total of 600 training images have been considered. Some examples of the training sequences are presented in Fig. 8.

Such images have been therefore classified by human operators into three classes: a) Too dark, b) Too bright and c) Good quality (just one for each sequence). The images classified as too dark or too bright have been used to compute the quality function BQ and the luminance value L. Then, the first NN has been trained with the following patterns:

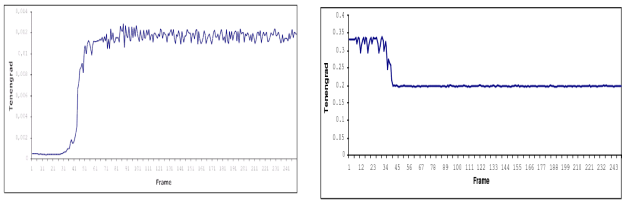
$$TS_{CB} = \{((BQ^i, L^i), Dark/Bright \,|\, i \in [1, .., 580]\}$$

where i is the index of the the current image and $Dark$ or $Bright$ its expected classification.

To train the following NNs, the images classified as too dark have been employed to train the opening NN. While the images too bright have been used to train the closing NN. Since the two procedures are similar, only the training of the opening NN is presented. For each image, the corresponding iris position $Iris$ and luminance value L_t have been inserted in the training pattern. In addition, the luminance value corresponding to the image related to the previous iris position L_{t-1} in the opening direction is also included in the training pattern. The expected value of the network is given by the number of iris positions (*steps*) between the current iris position and the iris position of the image classified as optimal.

V. Experimental Results

To test the effectiveness of the proposed method and its impact on active vision applications, the experiments have been first conducted on a IEEE-1394 SONY DFW-VL 500 camera mounted on a PTU platform. This set of experiments has been considered to verify the reliability and accuracy of the proposed method to focus on targets, to tune the target brightness and finally to adjust both parameters. A second round of experiments has been considered to compare the proposed solution to standard systems available on the market. This type of experiments aimed to demonstrate how the capability to tune parameters on selected image areas is of



(a) (b)

Fig. 9. Plot of the tenengrad value computed for the first 10 seconds of a sequence on the bounding box of the object of interest (a) and on the entire image (b).

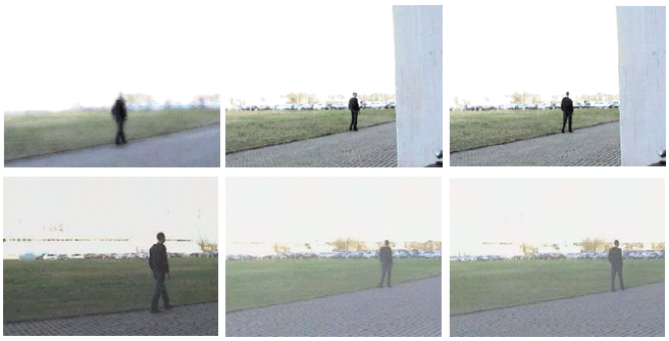
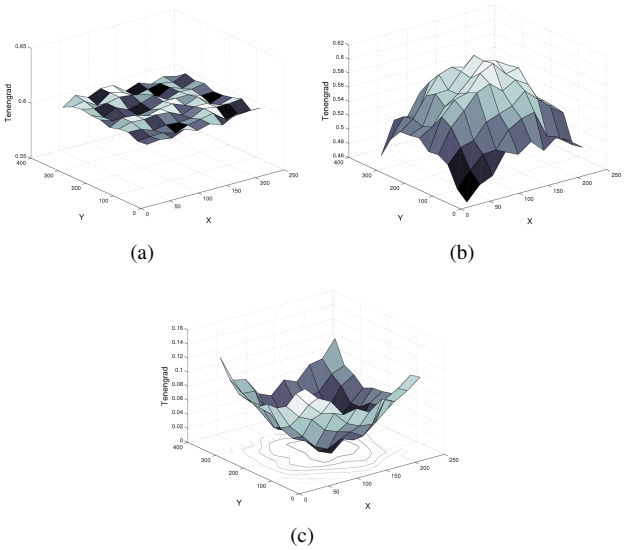


Fig. 10. Comparison between proposed strategy and automatic regulation. Top row regards the proposed tuning process the bottom row an automatic process.



(a) (b)

(c)

Fig. 11. Evaluation of the the proposed technique on the basis of the tenengrad value computed on a selected object with respect to its position inside the image. The chart (a) plots the tenengrad value computed on the object's bounding box when the proposed tuning strategy is applied. The chart (b) plots the tenengrad value computed on the object's bounding box when the auto focusing technique is adopted. Chart (c) plots the difference of the two approaches.

great advantage for applications based on image processing. To achieve such an objective, a standard Canon MV-600i handycam and an Axis PTZ-213 have been selected as metre to assess the proposed solution benefits. All the sequences used for the experimental evaluation have been independently acquired from those used during the training phase. As matter of fact, the chosen environment, the lighting and the weather conditions were different.

Focusing

The focusing capabilities of the proposed solution have been evaluated by tracking and focusing objects in an outdoor environment. In this context, the proposed method succeeded in focusing the objects in a period that spans from $0.5s$ to $2s$. This time range is in contrast with real-time applications. However, the transmission delays to control the camera and to acquire new images have to be taken into account in such an evaluation. A noticeable reduction of the focusing time could be achieved by implementing the proposed solution on-board to smart cameras.

The important thing is that, no matter which are the initial focus position and the defocusing degree of the objects, the hierarchy achieves its goal about focusing the object.

It is worth noticing how the proposed technique, by operating on information extracted only from the object of interest, introduces a side effect on the remaining areas of the image. In particular, by improving the acquisition quality of the object, the remaining regions of the image could exhibit a lower quality. This effect can be seen in Fig. 9 where the tenengrad values computed just within the bounding box of the object Fig. 9(a) and on the entire image Fig. 9(b) are plotted for the first 10 seconds of a test sequence. In these two charts the trend of the two curves is the opposite.

To compare the proposed solution with a commercial camera, a first experiment has been executed by shooting the same scene, represented by a moving object acquired in an outdoor environment, with both automatic and proposed regulation strategy. For both cases, during the acquisition process the tenengrad value has been computed on the bounding-box of the object of interest. Some frames of a test sequence can be seen in Fig. 10. The experiment has been conducted on 20 different sequences of $30s$ each. In this context, after few frames, needed to estimate the first *Optimum Focus* position, the performance of the proposed strategy shows a mean increment of the tenengrad value of about 15%. If we consider the case in which the object is not in the centre of the image, the increment is of about 25%. As can be seen in Fig. 11, when the position of the object of interest is not close to the centre of the image, the commercial approach shows a worsening of the tenengrad value (Fig. 11(b)). Instead, the proposed techinque, focusing on the object, maintains the tenengrad value almost constant regardless the position of the object inside the image (Fig. 11(a)). The difference of the two approaches is considerable (Fig. 11(c)). The overall gain of the proposed solution is estimated in a fair 6% for objects near the image center to a remarkable 23% when the objects are near to the corners of the image.

Brightness

Regarding the tuning of the image brightness the experiments have been performed by comparing the results obtained by the proposed and common techniques. In particular, the segmentation performance for both methods has been

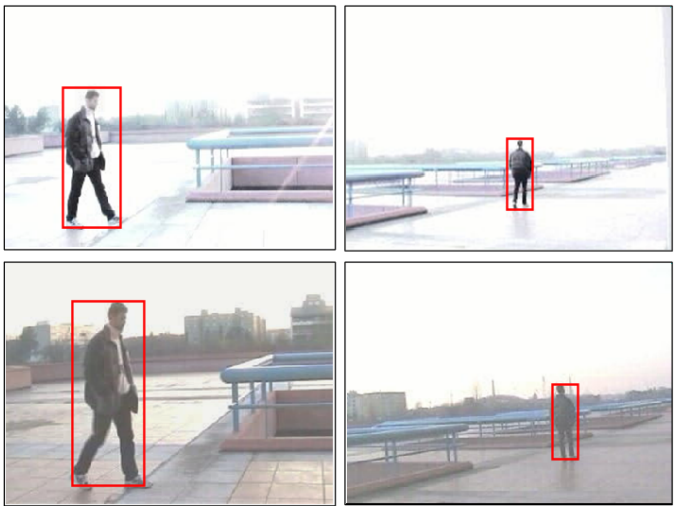


Fig. 12. Comparison between the proposed brightness tuning actuated on a Sony DFW-L 500 (top row) and an automatic regulation performed by a Canon V600-i (bottom row).

selected as evaluation metric. Due to the difficulties of creating the ground truth data of the blobs, the tests have been limited to 4 sequences of $30s$ each. A sample of the result obtained by the two approaches on a test sequence is shown in Fig. 12. It is worth noticing how the proposed solution generates images whose global quality is not good. They seem overexposed. But, if we consider just the region of the target, the contrast between the object and the background is much greater than in the common sequence.

Surveillance scenario

To evaluate the impact of the proposed strategy in the context of video surveillance applications, a last set of experiments has been taken into account. A first test shows the advantages of the proposed solution in tough lighting conditions. When the object moves in an area whose brightness is much different from the remaining part of the scene (shadow vs. sunlight), the proposed method shows an increment in the segmentation (i.e., number of corrected foreground pixels detected with respect to ground truth data) of about 35%. To show this, an Axis PTZ 213 with auotmatic or controlled iris regulation has been used. Fig. 13 shows some frames of the sequence acquired with the camera in autoiris mode. Fig. 14 shows some frames of the sequence acquired exploiting the proposed tuning strategy.

It is worth noticing that in the first case the automatic configuration is not able to handle the brightness compensation when the person walks in the bright area. As matter of fact, in the Fig.13(d),(i)-(l) frames, the upper part of the body is totally unnoticeable. In particular, in the frame Fig.13(k) the person has the arms wide open at the shoulder level. In these cases, there is no surveillance system able to segment the person to identify the silhouette for further analysis like behaviour understanding. To demonstrate this, in Fig. 13(h),(m)-(p) frames, the results obtained by the system described in [30] show how the torso of the person has not been detected.

On the opposite, in the same lighting conditions, the proposed strategy continuously tunes the iris parameter making it possible to acquire the person of interest with a good quality.

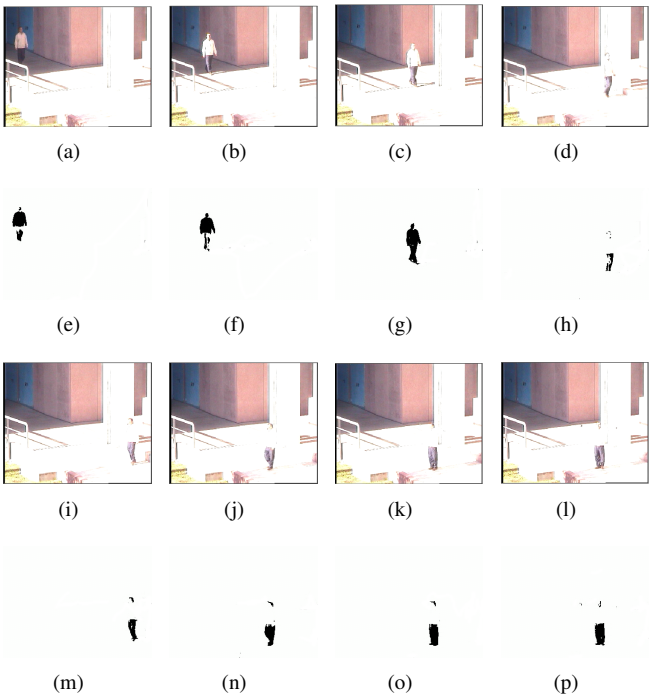


Fig. 13. Frames extracted from a sequence acquired with the camera Axis 213 in automatic reconfiguration mode. First and third rows present the acquired frames (full sequence available at $http : //users.dimi.uniud.it/ \sim christian.micheloni/TASE/AutomaticIndeo.rar.$). Second and fourth row show change detection images obtained by the system proposed in [30].

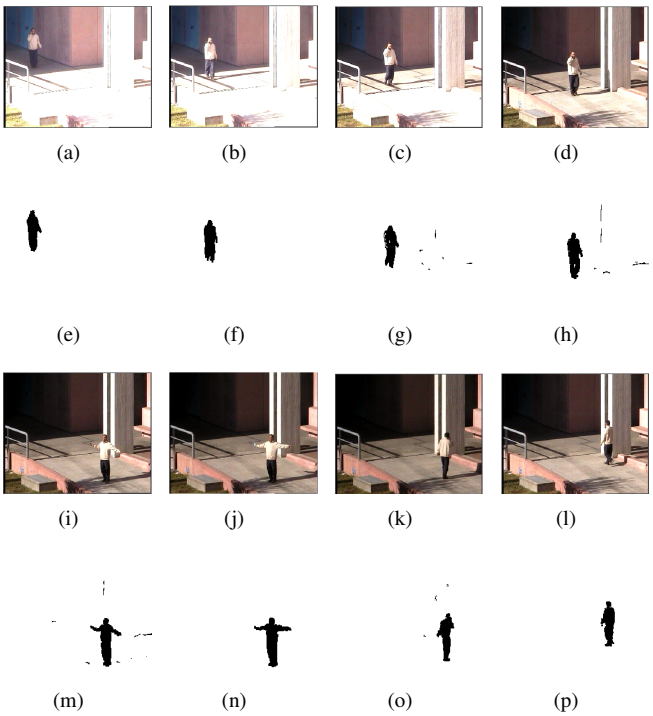


Fig. 14. Frames extracted from a sequence acquired with the camera Axis 213 using the proposed regulation strategy. First and third rows present the acquired frames (full sequence available at $http : //users.dimi.uniud.it/ \sim christian.micheloni/TASE/StrategyIndeo.rar.$). Second and fourth row show change detection images obtained by the system proposed in [30].

Indeed, comparing Fig. 14(j) with Fig. 13(k) it is clear how in similar conditions the acquisition quality of the proposed solution outperforms the acquisition quality of automatic techniques on-board of modern CCTV cameras. This difference is even more evident comparing the change detection results. All the frames 14(e)-14(h),14(m)-14(p) show how the same system [30] applied on images acquired by tuning the parameters returns the entire silhouette of the person. This result is of great importance for further processing steps.

A second test has been performed to show how the proposed technique could be coupled with usual active vision techniques. In particular, we have measured the accuracy of the blobs identification, computed by applying a change detection technique [30], on the images acquired by a moving camera. The proposed strategies (i.e. tuning both focus and iris) and a common camera approaches have been considered. Also in this context, the proposed method has shown good results by incrementing the blob segmentation of about 10% with peaks of 543%.

VI. Conclusions

In the current paper, a new method that aims to expand the paradigm of *Active Vision* has been proposed. In particular, the concept "*how to see the object*" has been introduced. This is a new way of thinking how a sensor should actively acquire an object of interest. To achieve such an objective, we have studied and developed a new method for enhancing the acquisition quality just on the image area related to the object of interest. A hierarchy of neural networks has been developed to control two main intrinsic camera parameters: Focus and Iris.

These two parameters are unequivocally related to the control of the image focus and brightness. It has been shown how the possibility to control these two parameters is relevant for improving surveillance applications. To achieve such results, quality operators have been exploited to define quality functions and tuning strategies. In particular, two quality functions have been used to classify the image and two operators have been used as principal values for the input patterns for the neural networks classifications.

The experimental results have shown how the proposed technique is really effective to control the quality of the image acquisition. Improving the quality of the acquired object improves the detection performance of low level techniques. This implies that new surveillance systems can be developed by considering the control of the acquisition quality as part of the loop for improving the system performance.

Acknowledgments

This work was partially supported by the Italian Ministry of University and Scientific Research within the framework of the project "Dynamic and unstructured environments interpretation, virtualization and monitoring by an integrated autonomous system of sensors and robots."(2009-2012)

References

[1] C. Regazzoni, V. Ramesh, and G. Foresti, "Special issue on video communications, processing, and understandingfor third generation surveillance systems," *Proceeding of the IEEE*, vol. 89, no. 10, Oct. 2001.

[2] Y. Y. L. Davis, R. Chellapa and Q. Zheng, "Visual surveillance and monitoring of human and vehicular activity," in *In Proceedings of DARPA97 Image Understanding Workshop*, 1997, pp. 19–27.

[3] G. Foresti, P. Mahonen, and C. Regazzoni, *Multimedia Video-Based Surveillance Systems: from User Requirementsto Research Solutions*. Kluwer Academic Publisher, Sep. 2000.

[4] G. Foresti, C. Micheloni, and C. Piciarelli, "Detecting moving people in video streams," *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2232–2243, 2005.

[5] P. Kumar, S. Ranganath, H. Weimin, and K. Sengupta, "Framework for real-time behavior interpretation from traffic video," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 43–43, Mar 2005.

[6] J. Zhou, D. Gao, and D. Zhang, "Moving vehicle detection for automatic traffic monitoring," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 1, pp. 51–59, Jan 2007.

[7] J. Ben-Arie, W. Zhiqian, P.Pandit, and S. Rajaram, "Human activity recognition using multidimensional indexing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1091–1104, Aug. 2002.

[8] S. Dockstader and A. Tekalp, "Multiple camera tracking of interacting and occluded human motion," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1441–1455, Oct. 2001.

[9] C. Piciarelli, C. Micheloni, and G. L. Foresti, "Trajectory-based anomalous event detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1544–1554, Nov. 2008.

[10] R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for cooperative multisensor surveillance," *Proceedings of the IEEE*, vol. 89, pp. 1456–1477, Oct. 2001.

[11] T. Gandhi and M. M. Trivedi, "Parametric ego-motion estimation for vehicle surround analysis usingan omnidirectional camera." *Machine Vision and Applications*, vol. 16, no. 2, pp. 85–95, 2005.

[12] T. Kanade, R. Collins, A. Lipton, P.Burt, and L.Wixson, "Advances in cooperative multisensor video surveillance," in *In Proceedings of DARPA Image Understanding Workshop*, vol. 1, Nov. 1998, pp. 3–24.

[13] A. Davison, I. D. Reid, and D. Murray, "The active camera as a projective pointing device," in *Proceedings of* 6^{th} *British Machine Vision Conference*, Birmingham UK, Sep.11-14 1999, pp. 11–14.

[14] D. Murray and A. Basu, "Motion tracking with an active camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 5, pp. 449–454, May 1994.

[15] M. Irani and P. Anandan, "A unified approach to moving object detection in 2d and 3d scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 6, pp. 577–589, June 1998.

[16] S. Araki, T. Matsuoka, N. Yokoya, and H. Takemura, "Real-time tracking of multiple moving object contours in a movingcamera image sequences," *IEICE Transaction on Information and Systems*, vol. E83-D, no. 7, pp. 1583–1591, July 2000.

[17] Y. Aloimonos, *Active Perception*, Y. Aloimonos, Ed. Lawrence Erlbaum Associates, 1993.

[18] F. Russo, "Automatic enhancement of noisy images using objective evaluation of image quality," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 4, pp. 1600–1606, 2005.

[19] G. Lighartand and F. Groen, "Comparison of different autofocus algorithms." in *International Conference on Pattern Recognition*, 1982, p. 597 600.

[20] M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques," *Journal of Optical Engineering*, vol. 32, pp. 2824–2836, 1993.

[21] M. Subbarao and J. Tyan, "The optimal measure for passive autofocusing and depth-from-focus," in *Symposium on Videometrics IV*, Philadelphia, 1995, pp. 89–99.

[22] F. Li and H. Jin, "A fast auto focusing method for digital still camera," in *International Conference on Machine Learning and Cybernetics*, vol. 8, Guangzhou, China, Aug. 18-21 2005, pp. 5001–5005.

[23] Q. Feng, K. Han, and X. chang Zhu, "An auto-focusing method for different object distance situation," *International Journal of Computer Science and Network Security*, vol. 7, no. 6, pp. 31–35, Jun 2007.

[24] E. Krotkov, "Focusing," *International Journal Computer Vision*, vol. 1, pp. 223–237, 1987.

[25] F. C. Groen, I. T. Young, and G. Ligthart, "A comparison of different focus functions for use in autofocus algorithms." *Cytometry*, vol. 6, pp. 81–91, Mar 1985.

[26] N. K. C. Nathaniel, P. A. N. Aun, and H. M. Ang, "A practical issues in pixel-based auto-focusing for machine vision," in *IEEE International Conference on Robotics & Automation*, Seoul, Korea, May 2001, pp. 2791–2796.
[27] V. Murino, G. Foresti, and C. Regazzoni, "A distributed probabilistic system fo adaptive regulation of image processing parameters," *IEEE Trans. Syst., Man, Cybern.*, vol. 26, no. 1, pp. 1–20, Jan. 1996.
[28] E.P.Krotkov, *Active Computer Vision by Cooperative Focus and Stereo*. Springer Verlag, 1989.
[29] N. Kehtarnavaz and H. J. Oh, "Development and real-time implementation of a rule-based auto-focus algorithm," *Real-Time Imaging*, vol. 9, pp. 197–203, 2003.
[30] C. Micheloni and G. Foresti, "Real time image processing for active monitoring of wide areas"," *Journal of Visual Communication and Image Representation*, vol. 17, no. 3, pp. 589–604, June 2006.
[31] G. Foresti and C. Micheloni, "Genrealized neural trees for pattern recognition," *IEEE Transaction on Neural Networks*, vol. 13, no. 6, pp. 1540–1547, Nov 2002.
[32] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Graphical Models and Image Processing*, vol. 29, pp. 273–285, 1985.

Gian Luca Foresti received the laurea degree cum Laude in Electronic Engineering and the Ph.D. in Computer Science from University of Genoa, Italy, in 1990 and in 1994, respectively. Since 2000 he is Professor of Computer Science at the Department of Mathematics and Computer Science (DIMI), University of Udine, where he is also Director of the Artificial Vision and Real-Time Systems (AVIRES) Lab. His main interests involve (a) computer vision and image processing, (b) multisensor data and information fusion, (c) pattern recognition and neural networks. Prof. Foresti is author or co-author of more than 200 papers published in International Journals and Refereed International Conferences and he has contributed in several books in his area of interest. He has been Guest Editor of a Special Issue of the Proceedings of IEEE on "Video Communications, Processing and Understanding for Third Generation Surveillance Systems" and of a Special Issue of the IEEE Transactions on Systems, Man and Cybernetics on "Ambient Intelligence". He is member of the International Association of Pattern Recognition (IAPR) and Senior member of IEEE.

Christian Micheloni (M.Sc.02, Ph.D. 06) received the Laurea degree (cum Laude) as well as a Ph.D. in Computer Science respectively in 2002 and 2006 from the University of Udine, Udine, Italy. He is assistant professor at the University of Udine. Since 2000 he has taken part to European research. He has co-authored more than 40 scientific works published in International Journals and Refereed International Conferences. He serves as a reviewer for several International Journals and Conferences. Dr. Micheloni's main interests involve active vision for scene understanding by images acquired by moving cameras, neural networks for the classification and recognition of the objects moving within the scene. He is also interested in pattern recognition techniques for both the automatic tuning of the camera parameters for an improved image acquisition and for the detection of the faces. All these techniques are mainly developed and applied for video surveillance purposes. He is member of of the International Association of Pattern Recognition (IAPR) and member of the IEEE.