PRACTICAL OBJECT TRACKING SYSTEM ON FPGA

Spirov Rosen Petrov

Technical University of Varna
1, Studentska str.
9010, Varna, Bulgaria
Tel.+35952383488; E-mail: rosexel@abv.bg.

Absrtract

In this thesis the autour proposed the hardware application of algorithm based on a new representation of the objectes with the graph pyramids. The objectes are examined at different, increasing, resolutions until it is possible to separate the multiple region in the objectes belonging to the occlusion. The graph matching algorithm is performed. For describing its behavior use the VHDL language and the Altera's Quartus tools to synthesize and Altera DE2 board to implement a simple hardware FPGA design.

Keywords: Image processing, object Détection, FPGA, VHDL

1. INTRODUCTION

One of the most interesting applications is that of the video surveillance systems [1]. Video surveillance systems are employed in airports, etc. where cameras are installed are installed to control docks, areas customers, etc. There are two main categories of active sensors. The ones based on structured light triangulation and the laser range scanners. Sensors belonging in the first class project a light stripe on the scene and use a camera to view it in Fig. 1.

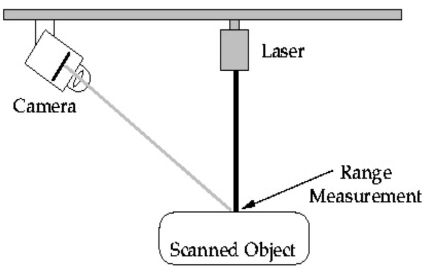

Fig. 1. The optical paths of structured light triangulation

The second category of sensors emit and receive a laser beam and by measuring difference of phase, time of flight or frequency, shift depth is measured and widely used in robotics applications.

A video analysis system has a modular structure. The first step of the processing is the segmentation of the scene, that is the separation of the frame in two set: the static part of the scene-background and the set of the moving objects of interests. This phase is called object detection. Detected moving objects are compared with moving objects in the previous frame to find correspondences. The trajectories of the objects, at point of the processing, are expressed in pixels. Pixel is marked as foreground if the inequality is satisfied:

$$| I_t (x, y) - B_t (x, y) | > T \qquad (1)$$

Where T is a predefined threshold. The background image B_t is updated by the use of a first order recursive filter as shown in:

$$B_{t+1} = \alpha I_t + (1 - \alpha) B_t a \qquad (2)$$

Where α is an adaptation coefficient. The basic idea is to provide the new incoming information into the current background image. Therefore transformation from pixel coordinates to real world coordinates is required. That is an Inverse Perspective Mapping Processing [2] is required and define two Euclidean Spaces: $W=\{(x,y,z)\} \in E^3$ is the 3-D space representing the real world; $I= \{(u,v)\} \in E^2$ is the 2-D space representing image in which the three-dimensional space is projected as image plane. The acquired image by camera belongs to the two-dimensional space I, while the transformed image by the IPM is defined with z=0 and belongs to the 3-D space W. The relationship between W and I is illustrated in Figure 2.

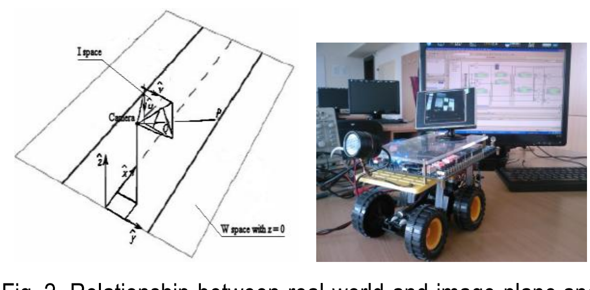

Fig. 2. Relationship between real world and image plane and camera model

2. THE SYSTEM TRACKING ALGORITHM

A number of objects is identified in each frame and a different label is associated to each object. Let $B^t = \{b_1^t,\ldots,b_n^t\}$ be the set of boxes belonging to the frame t, moreover let $L = \{l_1,\ldots,l_k\}$ be a set of labels. The element (i,j) of the matrix is 1 if the label assigned to the element b_j^t is the same as the label of b_i^{t-1}, it is 0 otherwise. Since there exists no duplicate label in set B^{t-1}, each row contains no more than one value set to 1. If the j-th row contains only zeros, it means that there exists no correspondence between an object belonging to the set B^{t-1} and b_j^t, i.e. b_j^t is a new box. If the j-th column contains only zeros, it means that there exists no correspondence between an object belonging to the set B^t and b_i^{t-1}.

As anticipated, the object tracking problem can be solved by computing a suitable injective mapping $\tau^t: B^t \rightarrow L$. This mapping solves a suitable Weighted Bipartite Graph Matching problem [3]. This is a graph where nodes can be divided into two sets such that no edge connects nodes in the same set. In our problem, the first set is B^{t-1}, while the second set is B^t. Before the correspondence is determined, each box of the set B^{t-1} is connected with each box of the set B^t, thus obtaining a Complete BG [4], as shown in Figure 3.

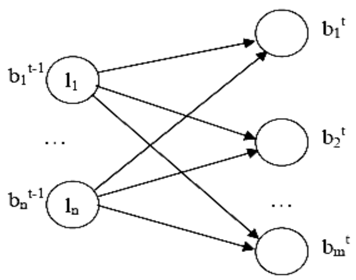

Fig. 3. The complete bipartite graph

Each box b_i^{t-1} is uniquely identified in the set B^{t-1} by its label. In general, an assignment between two sets B^{t-1} and B^t is any subset of $B^{t-1} \times B^t$, i.e., any set of ordered pairs whose first elements belongs to B^{t-1} and whose second elements belongs to Bt, with the constraint that each node may appear at most once in the set. A maximal assignment containing a maximal number of ordered pairs is known as a matching BGM [4].

The height *h* and the width *w* of each box is considered. The costelement (*i*, *j*) is the Euclidean distance between the (*h*, *w*) vectors of $b_i^{t-1}$and b_j^t. If the distance is greater than a threshold, the cost is ώ. Now, let $H_c^b(x)$ the histogram of the color *c* inside the box *b*. The correlation λ(i,j) index can compute as follows:

$$\lambda(i,j) = \sum_{c \in \{r,g,b\}} \frac{1}{3} \frac{\sum_{k=0}^{q-1} \min(H_c^{b_i^{t-1}}(k), H_c^{b_j^t}(k))}{\sum_{k=0}^{q-1} \max(H_c^{b_i^{t-1}}(k), H_c^{b_j^t}(k))} \quad (4)$$

Where λ (i,j) is the average of three indices, computed on the histograms of red, green and blue. The value of λ belongs to the interval [0,1]. The histogram of brightness Hb(x) is then computed, and its correlation λ is obtained. The cost function is computed as 1/ λ. In the next step let Gc(k) is the chromaticity histogram of the color c of a frame. The contribution of a color c λ {r, g, b} of a pixel to Gc(k) is given:

$$\lambda(i,j) = \frac{1}{3} \sum_{c \in \{r,g,b\}} \sum_{k=0}^{q-1} (G_c^{b_i^{t-1}}(k) - G_c^{b_j^t}(k)) \quad (5)$$

to the histogram is an integer valuebelonging to (0,…,*q*-l). We compute the correlation index λ (i,j) as follows:

$$\lambda_{tot} = \alpha \cdot \lambda_{metric1} + (1-\alpha) \cdot \lambda_{metric2} \quad (6)$$

Our method is based on a more complex representation, based on a graph pyramid, which in absence of occlusions retains the simplicity and effectiveness of the bounding box, but enables a more accurate object matching to take place during an occlusion. At lowest level there is an adjacency graph, where the nodes represent single pixels, and the edges encode the 4-connected adjacency relation. The intermediate levels are obtained by a bottom–up process, using the classical decimation-grouping procedure described in [5], where a colour similarity is used to decide which nodes must be merged. We decide to use the absolute difference between the mean R, G and B levels of the regions. For two regions R1and R2 the colour similarity is defined as follows:

$$C = \left|\mu_1^R - \mu_2^R\right| + \left|\mu_1^G - \mu_2^G\right| + \left|\mu_1^B - \mu_2^B\right| \quad (7)$$

Where μ_1^R,μ_2^R are the mean of R for regions R1 and R2, μ_1^G, μ_2^G are the mean of G for regions R1 and R2 μ_1^B, μ_2^B are the mean of B for regions R1 and R2.The representation differs from the quadtree structure [5]. Each intermediate node represents a sub region of the whole region, and its attributes are the position and size of the sub region bounding box and average colour of that sub region. In particular a node attribute is a set of 7 numbers {*w, h,* avR*,* avG*,* avB*,* cx*,* cy}

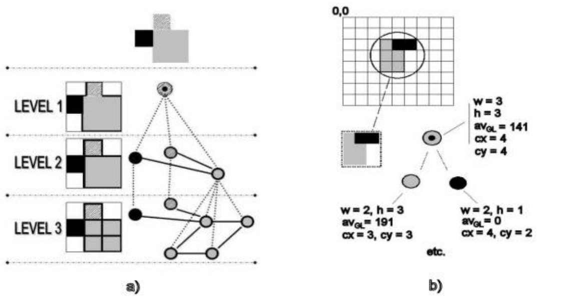

Fig. 4. The graph pyramid: a) the hierarchical structure; b) the nodes attributes

({*w, h,* avGL*,* cx*,* cy}) where *w*, *h* are the width and height of the sub region bounding box, avR, avG, avB (avGL) are the average colour/grey level of the sub region, cx, cy are the position of the sub region bounding box centre in the image on Figure 4. Tracking of the object is based on the features, requires selecting the right features, which plays a critical role in tracking.

In image processing, the RGB colour space is usually used to represent colour. Object boundaries usually generate strong changes in image intensities [6]. Edge detection is used to identify these changes. An important property of edges is that they are less sensitive to illumination changes compared to colour features. The Center of mass is vector of 1-by-n dimensions in length that specifies the center point of a region. For each point it is worth mentioning that the first element of the centroid is the horizontal x-coordinate of the center of mass, and the second element is the vertical y-coordinate [4,6] as in fig. 5.

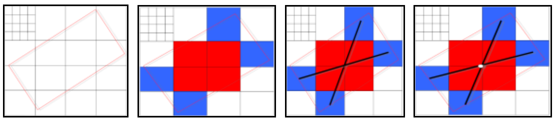
Fig. 5. The center point of a region

Texture is used for classification as well as tracking purpose. This feature is used to identify region or object in which we are interested. It is a measurement of the intensity variation of a surface which quantifies properties such as smoothness and regularity [6].


Fig. 6. The tracking object and system overview

In this method we compute the distance between the centroid that is smaller than a predefined threshold T. Suppose two objects Oc and Op, c for current frame and p for previous frame with center of mass (xc, yc) and (xp, yp) respectively,then the Euclidian distance between centers expressed as shown in equation:

$$\sqrt{(x_c - x_p)^2 + (y_c - y_p)^2} < T \qquad (8)$$

There are varies number of objects blobs in the current and previous frame I_n and $I_{n-1.}$ Let I_{n-1} and I_n be the number of objects in these frames.

The board Altera DE2 has an Analog Devices ADV7181 TV decoder chip which can convert NTSC composite video from camera to a digital format. Object recognition will be accomplished through software on the Nios II. The processes capture shown in fig. 7.

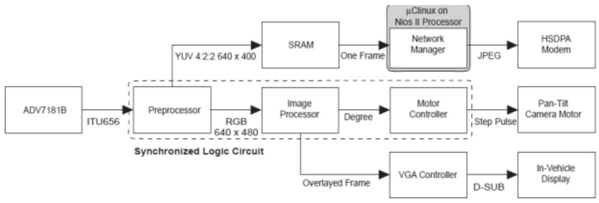

Fig. 7. The image processes capture

The ADV7181 decoder controller takes digital video input from the ADV each pixel from YUV to 16-bit RGB.A double line buffer in the FPGA's block RAM is used for data transfer between the 27 MHz frequency domain of the video controller and the 50 MHz frequency domain of the Avalon bus [7].

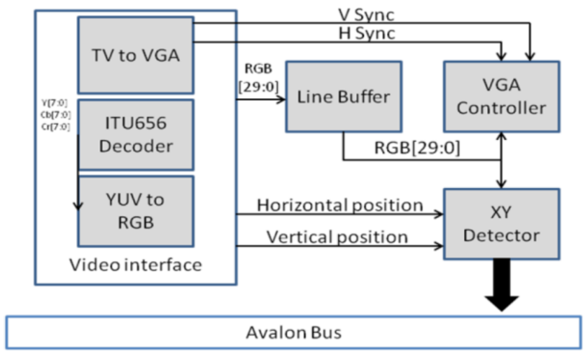

Fig. 8. Video interface and Avalon bus

The Nios II then performs processing on the buffer in SDRAM in order to find the center of the object we are tracking and to mark up the image.

The image processing software analyzes an image, termines the orientation of the target compared to the center of the screen, and then issues a command to turn either right or left to the Platform until the tracked object is in the center of the camera's field of view.

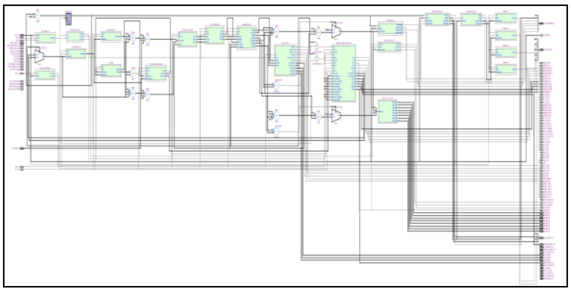

Fig. 9. The RTL project of object tracking system

3. CONCLUSION

In this thesis we propose an algorithm based on a new representation of the objects: the graph pyramids. This representation allows the resolutions of occlusions also in complex cases. The objects are examined at different, increasing, resolutions until it is possible to separate the multiple regions in the objects belonging to the occlusion. Furthermore, to detect the identities of the objects in the current frame beginning from the identities of the objects in the previous frame, a graph matching algorithm is performed.

References

[1] M. Bertozzi, A. Fascioli, "Vision-based intelligent vehicles: State of art and perspectives", Robotics and Autonomous Systems2000

[2] Poupard M, "Road edge traking using filtering theory" Palaiseau Ecole 2000.

[3] R. Sedgewick, Algorithms in C. Pt. 5, Graph algorithms. 3rd Edition Harlow : Addison-Wesley 2002, p. 482.

[4] Kuhn, H.W "The Hungarian Method for the Assignment Problem", Naval Research Logistics Quarterly, 2002

[5] J. M. Jolion, A. Montanvert. "The Adaptive Pyramid: A Framework for 2D Image Analysis". CVGIP: Image Under- standing, Academic Press. Vol. 55 – 3, pp. 339 – 348. 1992.

[6] H. Buxton and S. G. Gong, "Visual surveillance in a dynamic and uncertain world," Artif. Intell., vol. 78, Oct.1995

[7] www.altera.com