

A Bayesian Approach for Segmentation in Stereo Image Sequences

George A. Triantafyllidis

Information Processing Laboratory, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Thessaloniki 54006, Greece
Email: gatrian@iti.gr

Dimitrios Tzovaras

Informatics and Telematics Institute, 1st Km Thermi-Panorama Road, Thermi-Thessaloniki 57001, Greece
Email: tzovaras@iti.gr

Michael G. Strintzis

Information Processing Laboratory, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Thessaloniki 54006, Greece
Informatics and Telematics Institute, 1st Km Thermi-Panorama Road, Thermi-Thessaloniki 57001, Greece
Email: strintzi@eng.auth.gr

Received 31 August 2001 and in revised form 14 May 2002

Stereoscopic image sequence processing has been the focus of considerable attention in recent literature for videoconference applications. A novel Bayesian scheme is proposed in this paper, for the segmentation of a noisy stereoscopic image sequence. More specifically, occlusions and visible foreground and background regions are detected between the left and the right frame while the uncovered-background areas are identified between two successive frames of the sequence. Combined hypotheses are used for the formulation of the Bayes decision rule which employs a single intensity-difference measurement at each pixel. Experimental results illustrating the performance of the proposed technique are presented and evaluated in videoconference applications.

Keywords and phrases: Bayesian decision test, segmentation, stereoscopic video, disparity, motion.

1. INTRODUCTION

Stereo vision provides a direct way of inferring the depth information and therefore 3D perception by using two images (a stereo pair) destined for the left and right eye, respectively. Stereoscopic video sequences are represented by a number of consecutive stereo pairs each of which corresponds to a constant time instant. In a stereoscopic image sequence, each different view in a stereo pair is recorded with a difference in the observation angle, creating an enhanced 3D feeling to the observer and increased *tele-presence*, for example, in teleconferencing [1]. Such video can provide more vivid and accurate information about the scene structure than monoview video. For these reasons, stereoscopic video processing has been the focus of considerable attention in recent literature [2, 3, 4].

As stereo vision is used in more and more applications, efficiency in stereo sequence coding, transmission and storage becomes increasingly important. The major goal is to

exploit the correlation between the two frames (left and right) of a stereo sequence, along with the correlation which is present between the stereo pairs at consecutive time instances, in order to achieve reduction of the number of bits required to represent the stereo sequence [1, 4, 5]. Increased data compression will make the stereo imaging applications faster and more economical.

Accurate detection of occlusion and visible foreground and background areas in stereo is a valuable task for efficient coding of stereo imaging. Occlusion regions mark disparity discontinuity jumps which can be used to improve stereo image encoding and transmission, segmentation, motion analysis, and object identification processes. For example, it is highly useful for various stereoscopic coding methods [6], for sprite generation [7] and for background and depth extraction in stereoscopic videoconference sequences [8, 9].

Motion-compensated interframe coding of a sequence is one of the most effective techniques for reducing the quantity of transmitted information. This coding method predicts

a future frame from a previous frame by estimating the motion present in the scene and performing motion compensation on the previous frame [10]. The existence of the newly introduced pixels in an image frame degrades the quality of the motion compensation reconstructed image frames. This is due to the fact that the prediction method reduces prediction errors in static as well as moving regions but it fails where the scene background is uncovered by moving pixels since new picture information appears, for which no corresponding information can be found in the previous frame [11]. Therefore, an effective coding scheme should employ uncovered background pixel prediction, in addition to motion compensation for interframe coding.

The previous two paragraphs clearly imply that the detection of occlusion and visible foreground and background regions, along with the detection of uncovered-background areas is significant and valuable, since stereo image sequence processing requires information of the displacements created both by the motion of objects and also, by the disparity between two views of the 3D scene projects on two images.

These areas are defined as follows. Using only stereoscopic image information the scene may be segmented into the following areas [12, 13].

(i) The *occlusion areas* in one image of a stereo sequence, are spatially coherent groups of pixels in the background or the foreground that can be seen only in this frame and thus are occluded in the *other* counter frame of the same time instant.

(ii) The *visible foreground areas* in one image of a stereo sequence, consist of those pixels of the foreground, which are visible in both images of the same time instant.

(iii) The *visible background areas* in one image of a stereo sequence, are formed by those pixels of the background, which are visible in both images of the same time instant.

Correspondingly, in each stereo channel using only motion information the scene may be segmented into the following areas [14]:

(i) The *uncovered background areas* in one image of a stereo sequence, are formed by those pixels in the background, which are newly introduced in this image compared to the previous frame.

(ii) The *already existing areas* in one image of a stereo sequence, are groups of pixels that are displaced or are the same as in the previous frame.

We indicate as foreground the regions of the primary object having small values of depth, while background is the regions with large values of depth. Also note, at this point, that each pixel of one image of the stereo sequence can be classified in one of the first three regions as well as in one of the two last regions. Therefore, there exist *six* different classes—situations to which a pixel may belong.

The segmentation into visible and occlusion regions is treated as a secondary process in early stereoscopic research, postponed until matching is completed and smoothing is underway [15]. Techniques are also proposed that indirectly address the occlusion problem by minimizing spurious mismatches resulting from occlusion regions and discontinuities [16]. In [17], Belhumeur and Mumford point

out that occlusion areas must be identified and incorporated into matching process and, using Bayesian reasoning, they derive an energy functional using pixel intensity as the matching feature. Geiger et al. [18] also directly address occlusion regions by defining an a priori probability for the disparity field based upon a smoothness function and an occlusion constraint. Cox et al. [19] propose a dynamic programming solution to stereo matching based on matching edge-delimited intervals between corresponding scan lines. A similar strategy is used in [20, 21], where a dynamic programming algorithm is developed for the detection of significant disparity changes and large occlusion areas in a stereoscopic image pair. In this approach no smoothness or interscan-line compatibility constraints are used. Two additional constraints, namely the extended continuity constraint and the disparity gradient limit, are considered in [22].

As far as the detection of uncovered background areas is concerned, most methods utilize an algorithm based on change detection [23], evaluating the difference signal between the present and the previous frame. However, these methods are very sensitive to noise. In addition, most of the uncovered background techniques described in the literature use change detection for initial image segmentation and then use information provided by motion estimation to further detect the uncovered background pixels [24].

In the present paper, a novel approach is introduced for segmenting a videoconference stereo image sequence, employing a Bayes decision criterion followed by a simple checking procedure for the elimination of possible outliers. The only assumptions made for the stereoscopic video are related to the disparity and motion vector fields: the disparity and motion vectors are assumed to be relatively small and to obey a truncated Gaussian and a Gaussian distribution, respectively. The main advantage of the proposed algorithm is that it segments the frame using both the disparity and motion information, resulting in detecting *simultaneously* the occlusions and visible foreground and background regions as well as the uncovered-background areas.

This paper is organized as follows. Section 2 describes the mathematical analysis underlying the concept of Bayes testing. In Section 3, the calculation of the pdfs needed for the Bayes decision test is presented. Experimental results given in Section 4 evaluate visually and quantitatively the performance of the proposed method. Finally, conclusions are drawn in Section 5.

2. BAYES DECISION TEST

At two subsequent time instances, the following four frames are presented:

$$\begin{aligned} I_{rt_1} &: \text{right frame at time instant } t_1, \\ I_{lt_2} &: \text{left frame at time instant } t_2, \\ I_{rt_2} &: \text{right frame at time instant } t_2. \end{aligned} \quad (1)$$

The right frame at time instant t_2 (I_{rt_2}) will be assumed to

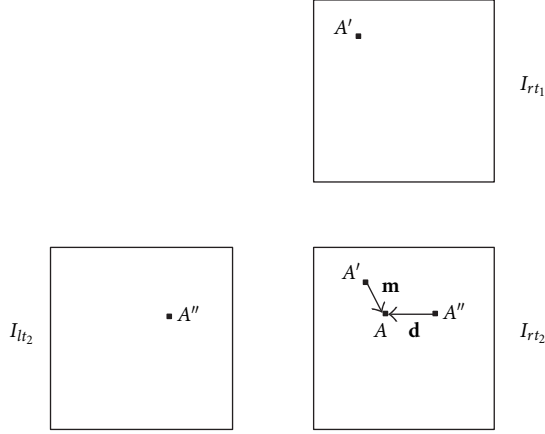


FIGURE 1: Disparity vector \mathbf{d} and motion vector \mathbf{m} defined for the frames I_{rt_1} , I_{lt_2} , I_{rt_2} .

be the current frame which will be segmented. Considering a point A at the current frame I_{rt_2} , the corresponding points at the frames I_{rt_1} and I_{lt_2} are the points A' and A'' , respectively. Then, the disparity vector \mathbf{d} and the motion vector \mathbf{m} are defined as shown in Figure 1.

The noisy intensity v_{lt_2} of each pixel of the left image of a stereo pair at time instant t_2 is given as follows.

Left image

$$v_{lt_2}(\mathbf{k}) = s_{lt_2}(\mathbf{k}) + w_{lt_2}(\mathbf{k}), \quad (2)$$

where the vector \mathbf{k} represents the spatial location $[x, y]^T$ of each pixel, $v_{lt_2}(\mathbf{k})$ is the noisy intensity value of the pixel located in $\mathbf{k} = [x, y]^T$ in the left image at time instant t_2 , $s_{lt_2}(\mathbf{k})$ is the respective noise-free intensity value in the left image at the same time instant, and $w_{lt_2}(\mathbf{k})$ is the zero-mean, additive, white Gaussian noise corrupting the left image.

Assuming the left I_{lt_2} to be the reference image, each pixel of the right image at time instant t_2 can be assigned to three different classes: the visible foreground ($A_{rt_2}^f$), the visible background ($A_{rt_2}^b$), and the occlusion ($A_{rt_2}^o$) classes of pixels, all defined in Section 1.

One of the most important parameters in the study of stereo vision is the disparity field. Assuming parallel axes camera configuration [25], two cameras with parallel optical axes, that is, image planes of the cameras are coplanar and collinear, and with identical optical characteristics, are employed to acquire the stereo pair sequence. In this case, the disparity \mathbf{d} reduces to the signed magnitude of the horizontal component d_x of the disparity vector, since the vertical component is always zero. Therefore, if the direction from the right image to the left image is considered as the positive direction, the disparities in stereo are always nonpositive if we regard the left image as reference and always nonnegative if we regard the right image as reference, due to camera geometry.

Then, the noisy intensity $v_{rt_2}(\mathbf{k})$ of the right image at the same time instant t_2 in terms of the noise-free intensity of the left image $s_{lt_2}(\mathbf{k})$ may be expressed as follows [13].

Right image

- if $\mathbf{k} \in A_{rt_2}^o$ then

$$v_{rt_2}(\mathbf{k}) = o_{rt_2}(\mathbf{k}) + w_{rt_2}(\mathbf{k}), \quad (3)$$

- if $\mathbf{k} \in A_{rt_2}^f$ then

$$v_{rt_2}(\mathbf{k}) = s_{lt_2}(\mathbf{k} - d_{x,l,t_2}^{(f)}(\mathbf{k})) + w_{rt_2}(\mathbf{k}), \quad (4)$$

- if $\mathbf{k} \in A_{rt_2}^b$ then

$$v_{rt_2}(\mathbf{k}) = s_{lt_2}(\mathbf{k} - d_{x,l,t_2}^{(b)}(\mathbf{k})) + w_{rt_2}(\mathbf{k}), \quad (5)$$

where $d_{x,l,t_2}^{(f)}(\mathbf{k})$ is the disparity of the foreground pixels at time instant t_2 , which is negative, since it is calculated in reference to the left image, $d_{x,l,t_2}^{(b)}(\mathbf{k})$ is the disparity of the background pixels at time instant t_2 , which is nonpositive, $o_{rt_2}(\mathbf{k})$ is the noise-free intensity of each occlusion pixel of the right image at t_2 and $w_{rt_2}(\mathbf{k})$ is the zero-mean, additive, white Gaussian noise corrupting the right image of the stereo pair. The terms $w_{lt_2}(\mathbf{k})$ and $w_{rt_2}(\mathbf{k})$ are assumed uncorrelated, since they express noises corrupting different images (the left and the right images of the stereo pair).

Assuming that the disparities $d_{x,l,t_2}^{(f)}(\mathbf{k})$ and $d_{x,l,t_2}^{(b)}(\mathbf{k})$ are small enough, a first order approximation of $s_{lt_2}(\mathbf{k} - d_{x,l,t_2}^{(f)}(\mathbf{k}))$ and $s_{lt_2}(\mathbf{k} - d_{x,l,t_2}^{(b)}(\mathbf{k}))$ may be used [14]. Then, (4) and (5) become

- if $\mathbf{k} \in A_{rt_2}^f$ then

$$v_{rt_2}(\mathbf{k}) = s_{lt_2}(\mathbf{k}) - g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k}) + w_{rt_2}(\mathbf{k}), \quad (6)$$

- if $\mathbf{k} \in A_{rt_2}^b$ then

$$v_{rt_2}(\mathbf{k}) = s_{lt_2}(\mathbf{k}) - g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(b)}(\mathbf{k}) + w_{rt_2}(\mathbf{k}), \quad (7)$$

where $g_{x,l,t_2}(\mathbf{k})$ is the horizontal component of the gradient vector of pixel \mathbf{k} in the left image at time instant t_2 .

We also define the differences

$$z_{(r-l),t_2}(\mathbf{k}) = v_{rt_2}(\mathbf{k}) - v_{lt_2}(\mathbf{k}), \quad (8)$$

$$w_{(r-l),t_2}(\mathbf{k}) = w_{rt_2}(\mathbf{k}) - w_{lt_2}(\mathbf{k}). \quad (9)$$

The above may be used to write the following equations:

- \mathbf{k} located in occlusion area, $A_{rt_2}^o$ area

$$z_{(r-l),t_2}(\mathbf{k}) = q_{(r-l),t_2}(\mathbf{k}) + w_{(r-l)}(\mathbf{k}), \quad (10)$$

- \mathbf{k} located in visible foreground area, $A_{rt_2}^f$ area

$$z_{(r-l),t_2}(\mathbf{k}) = -g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k}) + w_{(r-l),t_2}(\mathbf{k}), \quad (11)$$

- \mathbf{k} located in visible background area, $A_{rt_2}^b$ area

$$z_{(r-l),t_2}(\mathbf{k}) = -g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(b)}(\mathbf{k}) + w_{(r-l),t_2}(\mathbf{k}), \quad (12)$$

where $q_{(r-l),t_2}(\mathbf{k}) = o_{rt_2}(\mathbf{k}) - s_{lt_2}(\mathbf{k})$.

Now consider two consecutive right frames at time instances t_1 and t_2 . The noisy intensity of the first (previous) right frame at time t_1 is given by the following equation which is equivalent to (1).

Time instant t_1

$$v_{rt_1}(\mathbf{k}) = s_{rt_1}(\mathbf{k}) + w_{rt_1}(\mathbf{k}), \quad (13)$$

where $s_{rt_1}(\mathbf{k})$ is the noise-free intensity value in the right image at the same time instant and $w_{rt_1}(\mathbf{k})$ is the zero-mean, additive, white Gaussian noise corrupting the right image at time t_1 .

The noise free pixel intensity of the right frame at time instant t_2 , in terms of the pixel intensity of the right frame at time instant t_1 can be that of either an uncovered background value ($A_{rt_2}^u$), or an already existing value (displaced or stationary pixel) ($A_{rt_2}^e$). Then, the expression for the noisy intensity of the present frame (t_2) is as follows.

Time instant t_2

- if $\mathbf{k} \in A_{rt_2}^u$ then

$$v_{rt_2}(\mathbf{k}) = b_{rt_2}(\mathbf{k}) + w_{rt_2}(\mathbf{k}), \quad (14)$$

- if $\mathbf{k} \in A_{rt_2}^e$ then

$$v_{rt_2}(\mathbf{k}) = s_{rt_1}(\mathbf{k} - \mathbf{m}_{r,t_1}(\mathbf{k})) + w_{rt_2}(\mathbf{k}), \quad (15)$$

where $\mathbf{m}_{r,t_1}(\mathbf{k})$ is the nonuniform displacement vector due to motion, since it is calculated in reference to the right frame at time instant t_1 and $b_{rt_2}(\mathbf{k})$ is the noise-free intensity of each uncovered background pixel of the right image at t_2 .

Assuming that $\mathbf{m}_{r,t_1}(\mathbf{k})$ is small enough, a first-order approximation of $s_{rt_1}(\mathbf{k} - \mathbf{d}_{r,t_2}(\mathbf{k}))$ may be used, simplifying (15) to

- if $\mathbf{k} \in A_{rt_2}^e$ then

$$v_{rt_2}(\mathbf{k}) = s_{rt_1}(\mathbf{k}) - \mathbf{g}_{r,t_1}^T(\mathbf{k})\mathbf{m}_{r,t_1}(\mathbf{k}) + w_{rt_2}(\mathbf{k}), \quad (16)$$

where $\mathbf{g}_{r,t_1}(\mathbf{k})$ is the intensity gradient vector of the previous frame at \mathbf{k} .

In terms of the difference $y_{(t_2-t_1),r}(\mathbf{k}) = v_{rt_2}(\mathbf{k}) - v_{rt_1}(\mathbf{k})$ and (9) the following expressions are obtained:

- \mathbf{k} located in uncovered background area, $A_{rt_2}^u$ area

$$y_{(t_2-t_1),r}(\mathbf{k}) = p_{(t_2-t_1),r}(\mathbf{k}) + w_{(t_2-t_1),r}(\mathbf{k}), \quad (17)$$

- \mathbf{k} located in already existing area, $A_{rt_2}^e$ area

$$y_{(t_2-t_1),r}(\mathbf{k}) = -\mathbf{g}_{r,t_1}^T(\mathbf{k})\mathbf{m}_{r,t_1}(\mathbf{k}) + w_{(t_2-t_1),r}(\mathbf{k}), \quad (18)$$

where $p_{(t_2-t_1),r}(\mathbf{k}) = b_{rt_2}(\mathbf{k}) - s_{rt_1}(\mathbf{k})$.

Then, the six hypotheses H_i , $i = 0, \dots, 5$, are defined as follows:

$$\begin{aligned} H_0 & \text{ if } \mathbf{k} \in A_{rt_2}^o \text{ and } \mathbf{k} \in A_{rt_2}^u, \\ H_1 & \text{ if } \mathbf{k} \in A_{rt_2}^o \text{ and } \mathbf{k} \in A_{rt_2}^e, \\ H_2 & \text{ if } \mathbf{k} \in A_{rt_2}^f \text{ and } \mathbf{k} \in A_{rt_2}^u, \\ H_3 & \text{ if } \mathbf{k} \in A_{rt_2}^f \text{ and } \mathbf{k} \in A_{rt_2}^e, \\ H_4 & \text{ if } \mathbf{k} \in A_{rt_2}^b \text{ and } \mathbf{k} \in A_{rt_2}^u, \\ H_5 & \text{ if } \mathbf{k} \in A_{rt_2}^b \text{ and } \mathbf{k} \in A_{rt_2}^e. \end{aligned} \quad (19)$$

The Bayes decision test dictates that for each pixel \mathbf{k} , the hypothesis H_i is selected if

$$r_i(\mathbf{k}) < r_j(\mathbf{k}), \quad \forall j \neq i, \quad (20)$$

where $r_i(\mathbf{k})$ is the average cost of accepting the hypothesis H_i . This cost depends on the a priori probabilities $p(H_i)$ of hypotheses H_i , on the costs L_{mn} of deciding that H_m is correct while the actually correct hypothesis is H_n , and finally on the conditional probability density functions (pdfs) of $z_{(r-l),t_2}(\mathbf{k})$ and $y_{(t_2-t_1),r}(\mathbf{k})$. The average cost $r_i(\mathbf{k})$ is (see [26])

$$r_i(\mathbf{k}) = \sum_{j=0}^{M-1} L_{ij} p(H_j) f(z_{(r-l),t_2}(\mathbf{k}), y_{(t_2-t_1),r}(\mathbf{k}) | H_j), \quad (21)$$

where $M = 6$ in the case of (19).

Assuming that the a priori probabilities $p(H_i)$ and the costs L_{mn} are known, the conditional probability density functions $f(z_{(r-l),t_2}(\mathbf{k}), y_{(t_2-t_1),r}(\mathbf{k}) | H_j)$, must be calculated. The assumption of uncorrelated motion and disparity vectors for the right frame at time instant t_2 (see Figure 1), implies that

$$\begin{aligned} f(z_{(r-l),t_2}(\mathbf{k}), y_{(t_2-t_1),r}(\mathbf{k}) | H_j) \\ = f(z_{(r-l),t_2}(\mathbf{k}) | H_j) f(y_{(t_2-t_1),r}(\mathbf{k}) | H_j). \end{aligned} \quad (22)$$

Therefore, the probability density functions $f(z_{(r-l),t_2}(\mathbf{k}) | H_i)$ and $f(y_{(t_2-t_1),r}(\mathbf{k}) | H_i)$, $i = 0, \dots, 5$, must be computed in order to apply the Bayes decision test of (20).

3. CALCULATION OF THE PDFS

3.1. Calculation of $f(z_{(r-l),t_2}(\mathbf{k}) | H_i)$, $i = 0, 1$

Using the Bayes rule under the assumption that $\mathbf{k} \in A_{rt_2}^o$, we obtain

$$\begin{aligned} f(z_{(r-l),t_2}(\mathbf{k}) | H_i) &= \int_{\Omega_{q_{(r-l),t_2}}} f[z_{(r-l),t_2}(\mathbf{k}) | H_i, q_{(r-l),t_2}(\mathbf{k})] \\ &\quad \times f[q_{(r-l),t_2}(\mathbf{k})] d[q_{(r-l),t_2}(\mathbf{k})], \end{aligned} \quad (23)$$

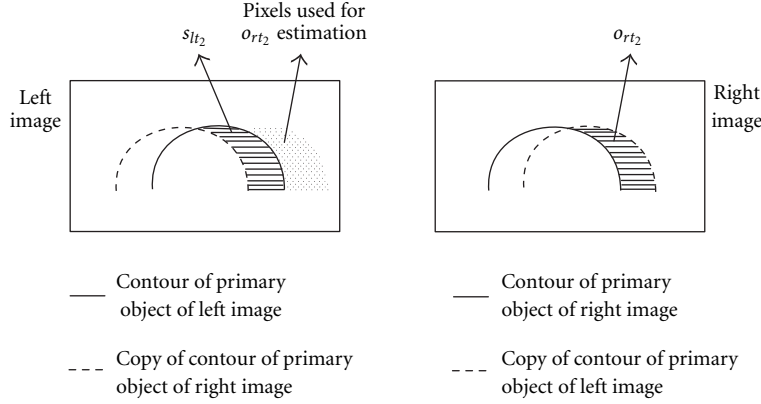


FIGURE 2: Regions used for the computation of $f[q_{(r-l),t_2}(\mathbf{k})]$.

where $\Omega_{q_{(r-l),t_2}}$ is the support range of $q_{(r-l),t_2}(\mathbf{k})$. Then, assuming independence for $z_{(r-l),t_2}(\mathbf{k})$ and $q_{(r-l),t_2}(\mathbf{k})$, trivially [26]

$$\begin{aligned} f[z_{(r-l),t_2}(\mathbf{k}) | H_i, q_{(r-l),t_2}(\mathbf{k})] \\ &= f_w[z_{(r-l),t_2}(\mathbf{k}) - q_{(r-l),t_2}(\mathbf{k})] \\ &= \frac{1}{\sigma_w \sqrt{2\pi}} \exp\left(-\frac{[z_{(r-l),t_2}(\mathbf{k}) - q_{(r-l),t_2}(\mathbf{k})]^2}{2\sigma_w^2}\right). \end{aligned} \quad (24)$$

However, in order to evaluate the integral of (23), knowledge of the density $f[q_{(r-l),t_2}(\mathbf{k})]$ is necessary. By definition, $q_{(r-l),t_2}(\mathbf{k})$ is the difference between the intensity of the occlusion areas $o_{rt_2}(\mathbf{k})$ of the right image and the intensity $s_{lt_2}(\mathbf{k})$ that the same pixels would have in the left image. Therefore, the region $s_{lt_2}(\mathbf{k})$ is a strip of the boundary pixels from inside the right part of the contour of the primary object of the left image [13]. We may also reasonably assume that the statistics of the adjacent region outside the right part of the contour of the primary object will be sufficiently similar to the statistics of the intensity of the occlusion areas $o_{rt_2}(\mathbf{k})$ (see Figure 2).

The intensity histogram ($f[s_{lt_2}(\mathbf{k})]$) of the right boundary pixels inside the primary object and the intensity histogram ($f[o_{rt_2}(\mathbf{k})]$) of the pixels outside the right boundaries of the primary object, may then be calculated, after defining the contour of the primary object in the left image, using the method of active contour models (or *snakes*) [27, 28]. Therefore, since $q_{(r-l),t_2}(\mathbf{k}) = o_{rt_2}(\mathbf{k}) - s_{lt_2}(\mathbf{k})$ and assuming independence for $o_{rt_2}(\mathbf{k})$ and $s_{lt_2}(\mathbf{k})$, the pdf of their difference will be (see [26])

$$f[q_{(r-l),t_2}(\mathbf{k})] = f[o_{rt_2}(\mathbf{k})] * f[s_{lt_2}(\mathbf{k})]^*, \quad (25)$$

where the star \star denotes the flipped histogram (since $q_{(r-l),t_2}(\mathbf{k}) = o_{rt_2}(\mathbf{k}) - s_{lt_2}(\mathbf{k})$) and the asterisk $*$ denotes convolution. Then, each pdf $f(z_{(r-l),t_2}(\mathbf{k}) | H_i)$, $i = 0, 1$, is easily derived from (23), (24), and (25).

3.2. Calculation of $f(z_{(r-l),t_2}(\mathbf{k}) | H_i)$, $i = 2, 3$

Applying the Bayes rule to $f(z_{(r-l),t_2}(\mathbf{k}) | H_i)$, $i = 2, 3$, under assumption that $\mathbf{k} \in A_{rt_2}^f$ we obtain [13]

$$\begin{aligned} f(z_{(r-l),t_2}(\mathbf{k}) | H_i) \\ &= \int_{-g_{x,l,t_2}(\mathbf{k})\Omega_{d_{x,l,t_2}^{(f)}}} f[z_{(r-l),t_2}(\mathbf{k}) | H_i, -g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k})] \\ &\quad \times f[-g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k})] d[-g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k})], \end{aligned} \quad (26)$$

where $\Omega_{d_{x,l,t_2}^{(f)}}$ is the support range of $d_{x,l,t_2}^{(f)}(\mathbf{k})$.

Since the difference $w_{(r-l),t_2}(\mathbf{k}) = w_{rt_2}(\mathbf{k}) - w_{lt_2}(\mathbf{k})$ is a zero-mean, additive, white Gaussian noise, the pdf of $w_{(r-l),t_2}(\mathbf{k})$ will be given by the following expression [14]:

$$f_w(w_{(r-l),t_2}(\mathbf{k})) = \frac{1}{\sigma_w \sqrt{2\pi}} \exp\left(-\frac{[w_{(r-l),t_2}(\mathbf{k})]^2}{2\sigma_w^2}\right), \quad (27)$$

where σ_w^2 is the noise variance of $w_{(r-l),t_2}(\mathbf{k})$, given by $\sigma_w^2 = \sigma_{w_{rt_2}}^2 + \sigma_{w_{lt_2}}^2$ since $w_{lt_2}(\mathbf{k})$ and $w_{rt_2}(\mathbf{k})$ were assumed uncorrelated.

Also recall that since $\mathbf{k} \in A_{rt_2}^f$, then the difference vector is

$$z_{(r-l),t_2}(\mathbf{k}) = -g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k}) + w_{(r-l),t_2}(\mathbf{k}). \quad (28)$$

Hence, if $d_{x,l,t_2}^{(f)}(\mathbf{k})$, $z_{(r-l),t_2}(\mathbf{k})$ are assumed to be independent random variables, then trivially [26]

$$\begin{aligned} f[z_{(r-l),t_2}(\mathbf{k}) | H_i, -g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k})] \\ &= f_w[z_{(r-l),t_2}(\mathbf{k}) + g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k})] \\ &= \frac{1}{\sigma_w \sqrt{2\pi}} \exp\left(-\frac{[z_{(r-l),t_2}(\mathbf{k}) + g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k})]^2}{2\sigma_w^2}\right). \end{aligned} \quad (29)$$

Due to the parallel camera geometry, since $\mathbf{k} \in A_{rt_2}^f$, the disparity $d_{x,l,t_2}^{(f)}(\mathbf{k})$ is negative

$$d_{x,l,t_2}^{(f)}(\mathbf{k}) < 0. \quad (30)$$

Then, the probability density function of $d_{x,l,t_2}^{(f)}(\mathbf{k})$ can be assumed to be a truncated Gaussian

$$f(d_{x,l,t_2}^{(f)}(\mathbf{k})) = \begin{cases} 0 & \text{if } d_{x,l,t_2}^{(f)}(\mathbf{k}) \geq 0, \\ \frac{2}{\sigma_{d_{x,l,t_2}^{(f)}} \sqrt{2\pi}} \exp\left(-\frac{[d_{x,l,t_2}^{(f)}(\mathbf{k})]^2}{2\sigma_{d_{x,l,t_2}^{(f)}}^2}\right) & \text{if } d_{x,l,t_2}^{(f)}(\mathbf{k}) < 0. \end{cases} \quad (31)$$

From the same assumption of uncorrelated $g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k})$ and $z_{(r-l),t_2}(\mathbf{k})$, using (28), it follows that

$$\sigma_{d_{x,l,t_2}^{(f)}}^2 = \frac{|\sigma_{z_{(r-l),t_2}}^2 - \sigma_w^2|}{g_{x,l,t_2}^2(\mathbf{k})}. \quad (32)$$

In this case, the pdf $f[-g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(f)}(\mathbf{k})]$ will also be a truncated Gaussian with variance $\sigma^2 = |\sigma_{z_{(r-l),t_2}}^2 - \sigma_w^2|$. Following [29], we will approximate $\sigma_{z_{(r-l),t_2}}^2$ by $z_{(r-l),t_2}^2(\mathbf{k})$. Then, each pdf $f(z_{(r-l),t_2}(\mathbf{k}) | H_i)$, where $i = 2, 3$, is easily determined by (26).

3.3. Calculation of $f(z_{(r-l),t_2}(\mathbf{k}) | H_i)$, $i = 4, 5$

In this case $\mathbf{k} \in A_{rt_2}^b$, precisely the same equations as in Section 3.2 for the case of foreground pixels may also be used, with obvious modifications, for the evaluation of $f(z_{(r-l),t_2}(\mathbf{k}) | H_i)$, $i = 3, 4$

$$f(z_{(r-l),t_2}(\mathbf{k}) | H_i) = \int_{-g_{x,l,t_2}(\mathbf{k})\Omega_{d_{x,l,t_2}^{(b)}}} f[z_{(r-l),t_2}(\mathbf{k}) | H_i, -g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(b)}(\mathbf{k})] \times f[-g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(b)}(\mathbf{k})] d[-g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(b)}(\mathbf{k})]. \quad (33)$$

However, very often in practice, either the background is sufficiently uniform so that $g_{x,l,t_2}(\mathbf{k}) \approx 0$ or far away from the camera (high depth) forcing the background disparity to very small values. In either such case $g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(b)}(\mathbf{k}) \approx 0$, hence the pdf is no longer conditioned on the value of a random number and reduces to the following expression:

$$f(z_{(r-l),t_2}(\mathbf{k}) | H_i) = \frac{1}{\sigma_w \sqrt{2\pi}} \exp\left(-\frac{[z_{(r-l),t_2}(\mathbf{k})]^2}{2\sigma_w^2}\right), \quad (34)$$

alternatively, this is also trivially seen to be the form of (33) reduces to, when $g_{x,l,t_2}(\mathbf{k})d_{x,l,t_2}^{(b)}(\mathbf{k}) \approx 0$.

3.4. Calculation of $f(y_{(t_2-t_1),r}(\mathbf{k}) | H_i)$, $i = 0, 2, 4$

In this case $\mathbf{k} \in A_{rt_2}^u$, an analysis similar to Section 3.1 is used. On applying the Bayes rule, we obtain [14]

$$f(z_{(t_2-t_1),r}(\mathbf{k}) | H_i) = \int_{\Omega_{q(t_2-t_1),r}} f[y_{(t_2-t_1),r}(\mathbf{k}) | H_i, p_{(t_2-t_1),r}(\mathbf{k})] \times f[p_{(t_2-t_1),r}(\mathbf{k})] d[p_{(t_2-t_1),r}(\mathbf{k})], \quad (35)$$

where $\Omega_{p_{(t_2-t_1),r}}$ is the support range of $p_{(t_2-t_1),r}(\mathbf{k})$. Then, if $y_{(t_2-t_1),r}(\mathbf{k})$ and $p_{(t_2-t_1),r}(\mathbf{k})$ are assumed to be independent random variables, trivially

$$\begin{aligned} f[y_{(t_2-t_1),r}(\mathbf{k}) | H_i, p_{(t_2-t_1),r}(\mathbf{k})] &= f_w[y_{(t_2-t_1),r}(\mathbf{k}) - p_{(t_2-t_1),r}(\mathbf{k})] \\ &= \frac{1}{\sigma_w \sqrt{2\pi}} \exp\left(-\frac{[y_{(t_2-t_1),r}(\mathbf{k}) - p_{(t_2-t_1),r}(\mathbf{k})]^2}{2\sigma_w^2}\right). \end{aligned} \quad (36)$$

From (35) and (36), it is clear that to evaluate the integral of (35) the pdf of $p_{(t_2-t_1),r}(\mathbf{k})$ must be computed. Recall that $p_{(t_2-t_1),r}(\mathbf{k}) = b_{rt_2}(\mathbf{k}) - s_{rt_1}(\mathbf{k})$, where $b_{rt_2}(\mathbf{k})$ is the intensity of the uncovered background and $s_{rt_1}(\mathbf{k})$ refers to the intensity of the scene pixels which have moved from the frame I_{t_1} to the frame I_{t_2} uncovering the background. Thus, $s_{rt_1}(\mathbf{k})$ is the intensity of the portions of the moving object which uncover background as the object moves. We assume that the statistics of the uncovered background $b_{rt_2}(\mathbf{k})$ is sufficiently described by the statistics of the neighboring scene background and the statistics of moving-object pixels $s_{rt_1}(\mathbf{k})$ are well approximated by the statistics of the primary object [14]. Using snakes for boundary detection as in Section 3.1 we obtain that

$$f[p_{(t_2-t_1),r}(\mathbf{k})] = f[b_{rt_2}(\mathbf{k})] * f[s_{rt_1}(\mathbf{k})]^*. \quad (37)$$

Then, using (36) and (37) each $f(y_{(t_2-t_1),r}(\mathbf{k}) | H_i)$ where $i = 0, 2, 4$, is easily obtained from (35).

3.5. Calculation of $f(y_{(t_2-t_1),r}(\mathbf{k}) | H_i)$, $i = 1, 3, 5$

Applying the Bayes rule to $f(y_{(t_2-t_1),r}(\mathbf{k}) | H_i)$, $i = 1, 3, 5$, under the assumption $\mathbf{k} \in A_{rt_2}^e$ we obtain [14]

$$f(y_{(t_2-t_1),r}(\mathbf{k}) | H_i) = \int_{\Omega_{\mathbf{m}_{r,t_1}}} f[y_{(t_2-t_1),r}(\mathbf{k}) | H_i, \mathbf{m}_{r,t_1}(\mathbf{k})] \times f[\mathbf{m}_{r,t_1}(\mathbf{k})] d[\mathbf{m}_{r,t_1}(\mathbf{k})], \quad (38)$$

where $f[\mathbf{m}_{r,t_1}(\mathbf{k})]$ and $\Omega_{\mathbf{m}_{r,t_1}}$ are the pdf and the domain of $\mathbf{m}_{r,t_1}(\mathbf{k})$. The motion vector for any \mathbf{k} is assumed to be Gaussian distributed [14] with zero-mean and covariance matrix $\mathbf{K}_{\mathbf{m}_{r,t_1}(\mathbf{k})}$, where

$$\mathbf{K}_{\mathbf{m}_{r,t_1}(\mathbf{k})} = \begin{bmatrix} \sigma_{m_x}^2(\mathbf{k}) & \sigma_{m_{xy}}(\mathbf{k}) \\ \sigma_{m_{xy}}(\mathbf{k}) & \sigma_{m_y}^2(\mathbf{k}) \end{bmatrix}, \quad (39)$$

in which $\sigma_{m_x}^2(\mathbf{k})$ and $\sigma_{m_y}^2(\mathbf{k})$ are the variances of $m_x(\mathbf{k})$ and

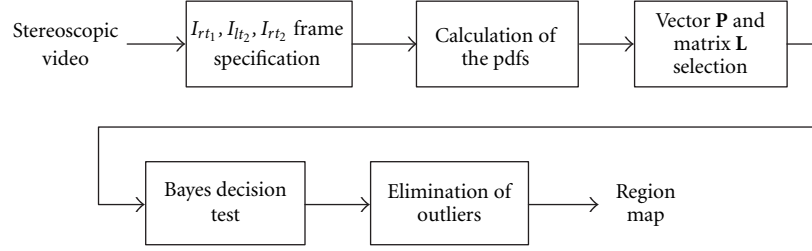


FIGURE 3: Flow graph of the proposed scheme.

$m_y(\mathbf{k})$, respectively, and $\sigma_{m_{xy}}(\mathbf{k})$ is the cross variance between $m_x(\mathbf{k})$ and $m_y(\mathbf{k})$. Assuming that the motion variance is the same in both x and y directions and that the $m_x(\mathbf{k})$ and $m_y(\mathbf{k})$ are uncorrelated with each other and with the noise, we find

$$\sigma_{m_x}^2(\mathbf{k}) = \sigma_{m_y}^2(\mathbf{k}) = \frac{|\sigma_y^2(\mathbf{k}) - \sigma_w^2(\mathbf{k})|}{\|\mathbf{g}_{r,t_1}(\mathbf{k})\|}, \quad (40)$$

where $\sigma_y^2(\mathbf{k})$ is approximated by $y_{(t_2-t_1),r}^2(\mathbf{k})$ [29].

Therefore, [14]

$$f(z_{(t_2-t_1),r}(\mathbf{k}) | H_i) = \frac{\exp\left(-\left(\frac{y_{(t_2-t_1),r}^2(\mathbf{k})/2\sigma_w^2} - (1/2)\mathbf{n}^T\mathbf{C}\mathbf{n}\right)\right)}{(2\pi)^{3/2}\sigma_w\|\mathbf{K}_{m,r,t_1}\|^{1/2}} I(y), \quad (41)$$

where

$$I(y) = \int_{\Omega_{m,r,t_1}} \exp\left\{-\frac{1}{2}(\mathbf{m} - \mathbf{n})^T\mathbf{C}(\mathbf{m} - \mathbf{n})\right\} d(\mathbf{m}), \quad (42)$$

$$\mathbf{n} = -\mathbf{C}^{-1} \frac{\mathbf{g}_{r,t_1}(\mathbf{k})}{\sigma_w^2} y_{(t_2-t_1),r}(\mathbf{k}),$$

$$\mathbf{C} = \frac{\mathbf{g}_{r,t_1}(\mathbf{k})\mathbf{g}_{r,t_1}(\mathbf{k})^T}{\sigma_w^2} + \mathbf{K}_{m,r,t_1}^{-1}.$$

4. EXPERIMENTAL RESULTS

This section presents the results of the simulation experiments using the proposed scheme. The algorithm was applied to both synthetic and natural stereo pair sequences to evaluate the performance of the proposed approach for segmenting stereo pair sequences into occlusion and visible background and foreground areas, along with the detection of the uncovered background areas.

The a priori probability vector \mathbf{P} will be defined by

$$\mathbf{P} = [p(H_0) \ p(H_1) \ p(H_2) \ p(H_3) \ p(H_4) \ p(H_5)], \quad (43)$$

where $p(H_i)$, $i = 0, \dots, 5$, is the a priori probability that hypothesis H_i is valid.

The cost matrix \mathbf{L} used in our experiments is the cost matrix \mathbf{L}' which corresponds to the maximum a posteriori probability—MAP test

$$\mathbf{L}'_{ij} = \begin{cases} 0 & \text{if } i = j, \\ 1 & \text{if } i \neq j. \end{cases} \quad (44)$$

The use of \mathbf{L}' assigns no cost to a correct decision and the same cost to all incorrect decisions.

Figure 3 describes the scheme which was used for the experiments. We use a noisy stereoscopic video as input and after the specification of the frames I_{rt_1} , I_{lt_2} , and I_{rt_2} , we calculate the pdfs needed for the Bayes decision test which follows. Then, after a simple post-processing step for outlier removal, we obtain the classification for the pixels into one of the six situations H_i , where $i = 0, \dots, 5$.

To resolve possible ambiguities in region classification a simple isolation technique is used, as a post processing step, in order to correct the classification of some pixels. This procedure produces a final re-classification into one of the six situations H_i , $i = 0, \dots, 5$. Specifically, a 4×4 neighborhood window is defined around the current pixel in the map produced by the Bayes decision test. If most of the pixels, belonging to this window, have been classified to other regions than that of the current pixel, then, the current pixel is re-classified to the most frequent situation H_i found in the window.

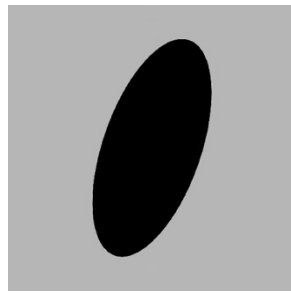
We first used the simple synthetic stereo pair sequence *Ellipse* which presents a uniform foreground and background, in order to evaluate the efficiency of the proposed algorithm. We employed the images I_{rt_1} , I_{lt_2} , I_{rt_2} as shown in Figure 4, which correspond, respectively, to the right frame at time instant t_1 , the left frame at time instant t_2 , and the right frame at time instant t_2 . Zero-mean, white Gaussian noises were added to both left and right channel with $\sigma_l = 7$ and $\sigma_r = 8$, respectively. We used the MAP cost matrix \mathbf{L}' and the a priori probability vector

$$\mathbf{P} = [0.1 \ 0.15 \ 0 \ 0.25 \ 0.15 \ 0.35]. \quad (45)$$

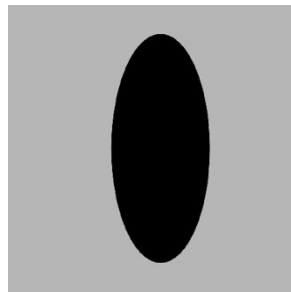
This is a typical selection for the vector \mathbf{P} for stereo video used for videoconference applications, which works well for such stereo videos as it will be shown next. Also, note that the probability $p(H_2)$ which corresponds to the probability of a pixel to belong to uncovered background and visible foreground areas is zero, since such a pixel classification cannot be true for these images.

Applying the proposed Bayes decision test followed by the checking procedure for the elimination of outliers, the segmented image of Figure 5 is obtained, where each greyscale colour corresponds to a different region as described from the hypotheses H_i .

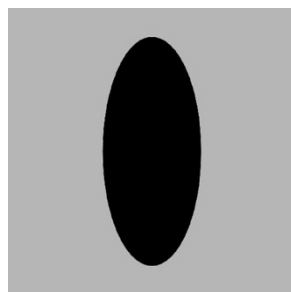
A similar synthetic stereo pair sequence is the *House* which is presented in Figure 6. This stereo sequence presents a foreground object with gradient colour, instead of solid



(a)

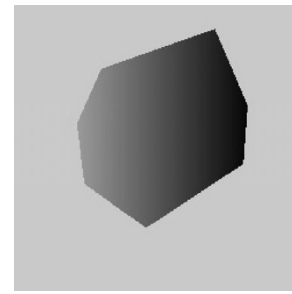


(b)

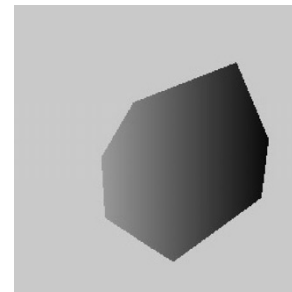


(c)

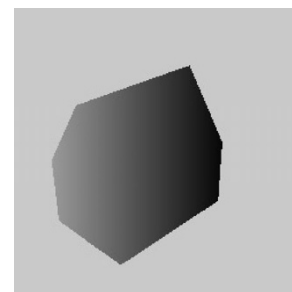
FIGURE 4: The generated test image sequence *Ellipse*, (a) frame I_{rt1} , (b) frame I_{lt2} , (c) frame I_{rt2} .



(a)



(b)



(c)

FIGURE 6: The generated test image sequence *House*, (a) frame I_{rt1} , (b) frame I_{lt2} , (c) frame I_{rt2} .

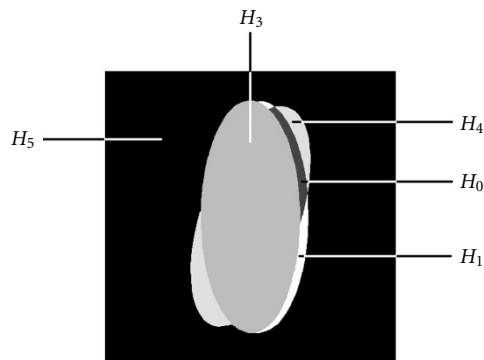


FIGURE 5: The segmentation map of *Ellipse*, produced from the proposed algorithm, indicating the regions which correspond to the situations H_i .

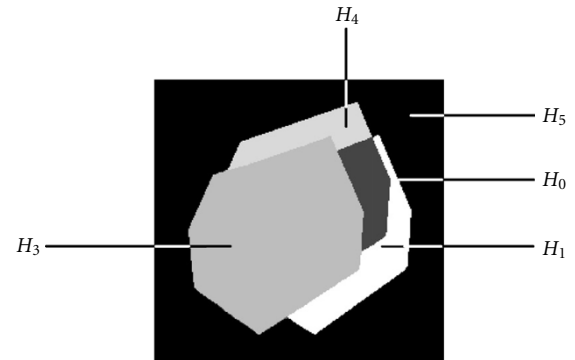


FIGURE 7: The segmentation map of *House*, produced from the proposed algorithm, indicating the regions which correspond to the situations H_i .

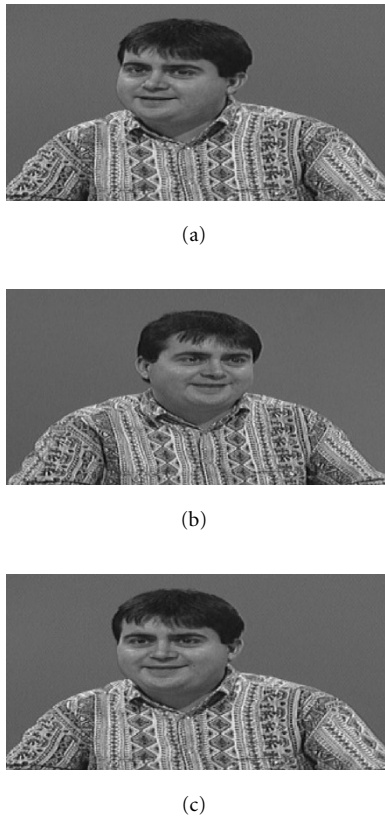


FIGURE 8: The stereoscopic real test image sequence *Ludo*, (a) frame I_{rt_1} , (b) frame I_{lt_2} , (c) frame I_{rt_2} .

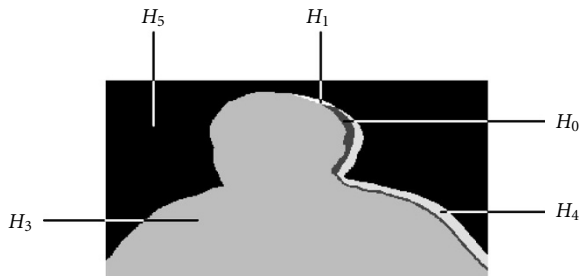


FIGURE 9: The segmentation map of *Ludo*, produced from the proposed algorithm, indicating the regions which correspond to the situations H_i .

colour as in *Ellipse*. Assuming the same vector \mathbf{P} and cost matrix \mathbf{L} as before, we obtain Figure 7 which illustrates the produced region map, indicating the pixel classification into one of the situations H_i , $i = 0, \dots, 5$.

To further evaluate the effectiveness of the proposed technique, results were obtained using the stereoscopic real image sequence *Ludo*.¹ Figure 8 shows the three images corresponding to the frames I_{rt_1} , I_{lt_2} , I_{rt_2} . After adding white Gaussian

¹This sequence was developed by Thomson Broadcast Systems for use in the PANORAMA (ACTS 092) EU project.



FIGURE 10: The stereoscopic real test image sequence *Sergio*, (a) frame I_{rt_1} , (b) frame I_{lt_2} , (c) frame I_{rt_2} .

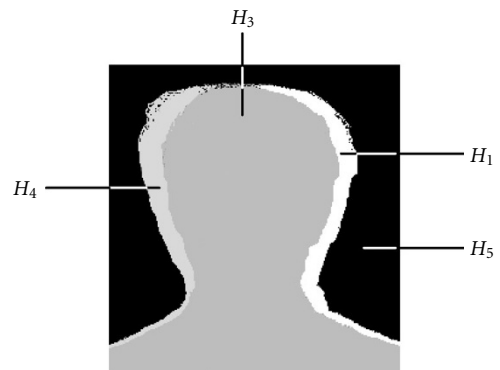


FIGURE 11: The segmentation map of *Sergio*, produced from the proposed algorithm, indicating the regions which correspond to the situations H_i .

noise to the images, the proposed Bayes decision test was applied, segmenting the frame I_{rt_2} . We assumed that the a priori probability vector \mathbf{P} is the same as in (45) and the MAP cost matrix \mathbf{L}' was also selected. Figure 9 illustrates the segmentation resulting from the application of the proposed algorithm.

Finally, the proposed algorithm was evaluated using the stereoscopic real image sequence *Sergio*. Figure 10 shows the three frames I_{rt_1} , I_{lt_2} , I_{rt_2} . The produced segmentation is presented in Figure 11. We must note at this point, that in this case, the produced region map does not have pixels belonging both to the uncovered background areas and the occlusion regions. This is due to the motion between the frames at the times t_1 and t_2 which does not uncover pixels that are occluded in the left frame at time t_2 .

5. CONCLUSIONS

In this paper, we introduce a novel Bayesian scheme for the segmentation of a videoconference noisy stereoscopic image sequence. A Bayes decision test is employed for detecting occlusions, visible foreground, and background regions and uncovered-background areas. Combined hypotheses are used for the formulation of the Bayes decision rule. It was assumed that the disparity and motion are small and obey a truncated Gaussian and a Gaussian distribution, respectively. The main advantage of the proposed method is that it segments the frame using both the disparity and motion information, resulting in detecting *simultaneously* the occlusions and visible foreground and background regions as well as the uncovered-background areas. Experimental results using both synthetic and natural stereo pairs sequences demonstrate the efficient performance of the proposed technique.

REFERENCES

- [1] M. Perkins, "Data compression of stereo pairs," *IEEE Trans. Communications*, vol. 40, no. 4, pp. 684–696, 1992.
- [2] N. Grammalidis, S. Malassiotis, D. Tzovaras, and M. G. Strintzis, "Stereo image sequence coding based on 3D motion estimation and compensation," *Signal Processing: Image Communication*, vol. 7, no. 2, pp. 129–145, 1995.
- [3] A. Tamtaoui and C. Labit, "Constrained disparity and motion estimators in 3DTV image sequence coding," *Signal Processing: Image Communication*, vol. 4, pp. 45–54, November 1991.
- [4] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-based coding of stereo image sequences using joint 3-D motion/disparity compensation," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 312–328, 1997.
- [5] M. G. Strintzis and S. Malassiotis, "Object-based coding of stereoscopic and 3D image sequences: a review," *IEEE Signal Processing Magazine*, vol. 16, no. 3, pp. 14–28, 1999, Special Issue on Stereo and 3D Imaging.
- [6] N. Grammalidis and M. G. Strintzis, "Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 8, no. 3, pp. 327–344, 1998.
- [7] N. Grammalidis, D. Beletsiotis, and M. G. Strintzis, "Multiview sprite generation and coding," in *Proc. IEEE International Conference on Image Processing*, Kobe, Japan, October 1999.
- [8] L. Falkenhagen, "Block-based depth estimation from image triples with unrestricted camera setup," in *Proc. IEEE Workshop on Multimedia Image Processing*, Princeton, NJ, USA, June 1997.
- [9] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "3-D camera motion estimation and foreground/background separation for stereoscopic image sequences," *Optical Engineering*, vol. 36, no. 2, pp. 574–579, 1997.
- [10] G. Tziritas and C. Labit, *Motion Analysis for Image Sequence Coding*, Elsevier, Amsterdam, 1994.
- [11] D. Hepper, "Efficiency analysis and application of uncovered background prediction in low bit rate image coder," *IEEE Trans. Communications*, vol. 38, no. 9, pp. 1578–1584, 1990.
- [12] G. A. Triantafyllidis, D. Tzovaras, and M. G. Strintzis, "Efficient classification of occlusion areas in stereo pairs," in *Proc. International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging: IWSNHC3I '99*, Santorini, Greece, September 1999.
- [13] G. A. Triantafyllidis, D. Tzovaras, and M. G. Strintzis, "Occlusion and visible background and foreground areas in stereo: a Bayesian approach," *IEEE CSVT*, vol. 10, no. 4, pp. 563–575, 2000, Special Issue on 3D Video Technology.
- [14] K. E. Matthews and N. M. Namazi, "A Bayes decision test for detecting uncovered-background and moving pixels in image sequence," *IEEE Trans. Image Processing*, vol. 7, no. 5, pp. 720–728, 1998.
- [15] U. R. Dhond and J. K. Aggarwal, "Structure from stereo: A review," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1489–1510, 1989.
- [16] C. Chang, S. Catterjee, and P. R. Kube, "On an analysis of static occlusion in stereo vision," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 722–723, Kauai, Hawaii, 1991.
- [17] P. Belhumeur and D. Mumford, "A Bayesian treatment of the stereo correspondence problem using half-occluded regions," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 506–512, Champaign-Urbana, Ill, USA, 1992.
- [18] D. Geiger, B. Ladendorf, and A. Yuille, "Occlusions and binocular stereo," in *Proc. European Conference on Computer Vision*, pp. 425–433, Santa Margherita Ligure, Italy, May 1992.
- [19] I. J. Cox, S. Hingorani, B. Maggs, and S. Rao, "Stereo without regularization," NEC Research Institute Report, NEC Research Institute, October 1992.
- [20] S. S. Intille and A. F. Bobick, "Disparity-space images and large occlusion stereo," Tech. Rep. 220, MIT Media Lab Perceptual Computing Group, 1993.
- [21] S. S. Intille and A. F. Bobick, "Incorporating edges in the recovery of occlusion region," Tech. Rep. 246, MIT Media Lab Perceptual Computing Group, 1994.
- [22] L. Falkenhagen, R. Koch, A. Kopernik, and M. G. Strintzis, "Disparity estimation based on 3-D arbitrarily shaped regions," Tech. Rep. R2045/UH/DS/P/023/b1 RACE Project R2045 Digital Stereoscopic Imaging and Applications (DISTIMA), 1994.
- [23] R. Thoma and M. Bierling, "Motion compensating interpolation considering covered and uncovered background," *Signal Processing: Image Communication*, vol. 1, no. 2, pp. 191–212, 1989.

- [24] C. Lettera and L. Masera, "Foreground/background segmentation in videotelephony," *Signal Processing: Image Communication*, vol. 1, no. 2, pp. 181–189, 1989.
- [25] H. Aydinoglu and M. Hayes III, "Stereo image coding: A projection approach," *IEEE Trans. Image Processing*, vol. 7, no. 4, pp. 506–516, 1998.
- [26] A. Papoulis, *Probability Random Variables and Stochastic Processes*, McGraw-Hill, New York, NY, USA, 3rd edition, 1991.
- [27] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," in *Proc. 1st International Conference on Computer Vision*, pp. 259–269, London, UK, 1987.
- [28] S. Malassiotis and M. G. Srintzis, "Tracking the left ventricle in echocardiographic images by learning heart dynamics," *IEEE Trans. on Medical Imaging*, vol. 18, no. 3, pp. 282–291, 1999.
- [29] N. M. Namazi, P. Penafiel, and C. M. Fan, "Nonuniform image motion estimation using the Kalman filtering," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 678–683, 1994.

George A. Triantafyllidis was born in Thessaloniki, Greece, in 1975. He received the Diploma degree from the Electrical Engineering Department of Aristotle University of Thessaloniki, Greece in 1997. He is currently associated with the Information Processing Laboratory of Aristotle University of Thessaloniki, where he is pursuing the Ph.D. degree, and the Informatics and Telematics Institute, Thessaloniki, Greece, as a Research Associate. He has participated in several research projects funded by EU and the Greek Secretariat of Research and Technology. Since 2000, he serves as a Teaching Assistant in Aristotle University. His research interests include image compression and analysis as well as monoscopic and stereoscopic image sequence coding and processing.



Dimitrios Tzovaras received the Diploma degree in electrical engineering and the Ph.D. degree in 2D and 3D image compression from Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1992 and 1997, respectively. He is now a Researcher in the Informatics and Telematics Institute of Thessaloniki. Prior to his current position, he was a Leading Researcher on 3D imaging at the Aristotle University of Thessaloniki. His main research interests include image compression, 3D data processing, virtual reality, medical image communication, 3D motion estimation, and stereo and multiview image sequence coding. His involvement with those research areas has led to the coauthoring of more than 20 papers in refereed journals and more than 50 papers in international conferences. He has served as a regular reviewer for a number of international journals and conferences. Since 1992, he has been involved in more than 20 projects in Greece, funded by the EC, and the Greek Ministry of Research and Technology. Dr. Tzovaras is a member of the Technical Chamber of Greece.



Michael G. Srintzis received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece in 1967, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, USA in 1969 and 1970, respectively. He then joined the Electrical Engineering Department at the University of Pittsburgh, Pittsburgh, Pa., where he served as Assistant (1970–1976) and Associate Professor (1976–1980). Since 1980 he is Professor of electrical and computer engineering at the University of Thessaloniki, and since 1999 Director of the Informatics and Telematics Research Institute in Thessaloniki, Greece. Since 1999 he serves as an Associate Editor of the *IEEE Trans. On Circuits and Systems for Video Technology*. His current research interests include 2D and 3D image coding, image processing, biomedical signal and image processing and DVD and Internet data authentication and copy protection. In 1984, Dr. Srintzis was awarded one of the Centennial Medals of the IEEE.

