# Spatiotemporal Spectral Coding of Stereo Image Sequences

Françoys Labonté, Chon Tam Le Dinh, Member, IEEE, Jocelyn Faubert. and Paul Cohen, Member, IEEE

Abstract—This paper presents a new compression scheme for interlaced stereoscopic sequences which differentiates between a region of fixation and a peripheral area, and thereby compacts the stereoscopic information into the spectral space of a monocular video channel. Spectral compression is achieved by avoiding transmitting high-frequency information over the entire images, but only within and around the region where the observer acuity is the highest. The proposed approach consists of decomposing the left and right fields of the stereoscopic pairs into low-pass and high-pass components. High-frequency components are then limited to a fixation region, thus allowing a reduction of their spectral extent. A composite video signal is then formed by positioning the different components into the available spectral space through filtering and modulation. The approach is compatible with the NTSC standard in the sense that the same color subcarrier and the same spectral region are used for the chrominance components. Strategies are proposed for the estimation of the fixation region, based upon a psychophysical study on visual strategies during depth discrimination tasks.

*Index* Terms—Fixation region, image coding, NTSC video signal, spectral compression, stereoscopy.

## I. INTRODUCTION

**S**TEREOSCOPIC imaging systems, which are increasingly used for multimedia, telepresence, or telerobotics, permit us to enhance tridimensional perception by presenting to each eye a slightly different viewpoint of the same scene. These systems have been shown to be beneficial for the accomplishment of a number of tasks, such as object traclung [1], object manipulation [2], and relative depth perception [3]. Since the usage of stereoscopic systems implies a significant increase in the amount of visual information for remote transmission, it is essential to make use of compact representations of this information. The representations should, on one hand, take advantage of all of the redundancy sources present in stereoscopic signals, and on the other hand, only exploit the information which is strictly necessary to the human observer.

Manuscript received June 13, 1998. This work was supported by the NSERC-Noranda Chair in Mining Automation, the Canadian Program of Networks of Centres of Excellence (IRIS), and by the FCAR funding agency from the government of Quebec. This paper was recommended by Associate Editor **S**. Panchanathan.

F. Labonté and P. Cohen are with the Perception and Robotics Laboratory, Ecole Polytechnique. Montreal. P.Q. H3C 3A7 Canada.

C. T. Le Dinh is with the Department of Electrical Engineering, Université de Sherbrooke, Sherbrooke, P.Q., Canada.

J. Faubert is with Ecole d'Optométrie, Université de Montreal, Montréal, P.Q. Canada.

Publisher Item Identitier S 1051-8215(99)02040-6.

# A. Stereoscopic Signals Characteristics

In general, image signals possess intraimage redundancy caused by the smoothness and cohesiveness of the objects forming a scene. Fixed image coding techniques, such as predictive or transform coding [4], exploit this intraimage redundancy by taking advantage of the fact that locally continuous and smooth surfaces will project onto image regions of homogeneous or smoothly varying brightness separated by discontinuities. In dynamic scenes, objects motion continuity generates temporal interimage redundancy whose elimination can lead to a significant reduction in the amount of information to be transmitted. Motion compensation represents the most popular coding strategy for dynamic images, and consists of transmitting a displacement vectors field which, applied to the previous image, allows the reconstruction of the current image [5]. One displacement vector per region of predefined shape and size is generally estimated in the original image. In addition to the displacement vectors, appropriately quantized prediction errors (i.e., the differences between the predicted and the real brightness values) can also be transmitted to improve the quality of the reconstructed images.

Stereoscopic signals possess an additional source of redundancy created by the observation of the same scene from two slightly different viewpoints. Most parts of a scene are thus visible from the two viewpoints, and project onto similar regions of the left and right image planes. In theory, this geometrical redundancy can be exploited for compression purposes. A possible strategy consists of putting into correspondence regions of preestablished size and shape of a stereoscopic pair, and transmitting disparity vectors allowing the reconstruction of one viewpoint from the other. This prediction strategy is referred to as *disparity compensation* [6], [7].

## B. Stereoscopic Perception Characteristics

Characteristics and limitations of the human binocular perception may be taken into account to compress the representations for stereoscopic information beyond the limits imposed by the theory of signal processing. For instance, depth resolution limits of the human visual system can be considered in order to limit disparity vectors amplitude to values greater than the depth discrimination threshold [8], [9]. Also, some stereoscopic image presentation systems may not reproduce the full disparity range perceivable by the human visual system [9]. The fact that the depth resolution of the human visual system decreases exponentially with an increase in distance from the fixation plane can also be exploited to represent disparity vectors with a nonuniform resolution according to an exponential quantization function [10]. Psychophysical data reported in [113 show that the threshold of stereoscopic perception as a function of increasing angular separation between a target and a fixation reference increases according to an exponential function, with maximal stereoacuity at the fovea and a more rapid stereoacuity decrease after 6° of visual angle [11]. Spatial acuity is also maximal at the fovea, and decreases as a monotonic function of the distance of target from fixation because of the decreasing density of cone receptors [12]. These facts can be taken into account to reduce the level of details in the peripheral regions [13].

The two main theories related to human stereoscopic vision, the suppression and the fusion theories [11], also play an important role in the elaboration of stereoscopic coding schemes. The suppression theory postulates that the subjective three-dimensional (3-D) percept resulting from a stereo pair with a degraded image possesses a quality level approaching the one of the nondegraded image [11]. The degraded image contributes to disparity computation, and is then suppressed to be replaced by the other image. The suppression theory can be considered as the basis of asymmetrical coding schemes, where the two images of a stereoscopic pair are treated differently, with one channel generally used as reference for the other channel prediction. Asymmetrical coding schemes often take for granted that the degradations resulting from the high compression of one of the two images of the pair are only slightly perceivable because the other image of the same pair keeps a good quality level [14], [15].

According to the fusion theory, the two images of a stereoscopic pair are fused to give rise to a volumetric representation of the environment [11]. This theory motivates symmetrical coding schemes where an intermediate representation integrating information from the two stereoscopic channels is created. With symmetrical approaches, reconstruction errors are generally uniformly distributed between the two viewpoints. The intermediate representation can take different forms, such as another image obtained by combining the two original images or a 3-D model integrating information from the two viewpoints.

## C. Review of Stereoscopic Compression Methods

Among the various compression methods for stereoscopic images, the simplest one consists of the independent coding of the left and right channels. However, this approach is far from efficient since the inherent stereoscopic redundancy is not exploited. Asymmetrical approaches, which exploit the suppression theory by subsampling one of the two images, have been used [14]–[16]. Even though these methods are attractive because they require no matching, they might not be adequate for applications requiring fine details perception [14], such as telemanipulation, for instance. Disparity compensation represents another example of asymmetrical coding [6], [7]. Other geometric relations than the ones concerning disparity can also be exploited. For instance, information about projective transformations to reconstruct one viewpoint from the other can be transmitted [17].

Symmetrical approaches which involve the creation of an intermediate representation can also be considered. For instance, an intermediate image corresponding to the image that would be seen from a camera positioned at the center of the stereoscopic system can be obtained from the disparity field between the two original stereoscopic images. Intensity values in this image are obtained by averaging the intensity of the corresponding pixels in the left and right images [18]. Another method to obtain an intermediate image consists of temporally multiplexing the left and right images of a stereoscopic sequence in an adaptive manner which takes into account the amplitude of the reconstruction errors coming from motion- or disparity-compensated predictions [19]. In order to maintain the consistency of the intermediate image, one of the two viewpoints must be shifted by an amount which depends upon disparity.

For the case of stereoscopic sequences, many redundancy sources are simultaneously available: intraimage structure, motion, and stereoscopy. Various methods, inspired from the MPEG compression standard, have been proposed to take into account the stereoscopic redundancy. A selection module is generally used for encoding image blocks by considering the reconstruction errors at the receiver. The left channel is coded by means of a discrete cosine transform (DCT) and motion compensation, and the right channel by motion compensation, disparity compensation, or direct block encoding with DCT, choosing for each block the estimate yielding the smallest reconstruction error [19], [20]. Disparity or motion compensation can also be applied to an intermediate image obtained by multiplexing the left and the shifted right viewpoints [19].

In order to avoid artifacts associated with block-based approaches when high compression rates are applied, it is possible to treat the image information in terms of twodimensional (2-D) objects [21]. An iterative procedure using the available stereoscopic information segments the images into uniform regions corresponding to objects. Motion information is also taken into account during the segmentation. Moving objects are then transmitted with a set of parameters describing their color, form, motion, and disparity. In regions which cannot be modeled satisfactorily, the temporal difference between successive images is transmitted in a conventional manner. Stereoscopic and dynamic information can also be integrated into a common and compact description of the scene, consisting of a 3-D structure and 3-D motion parameters [22]–[25]. Object modeling can facilitate motion and disparity estimation by talung into account, for instance, the difference between the current image and the projection of the estimated 3-D model [26]. The transmitted information can be represented in terms of 3-D objects constituted of planar surfaces, with motion described by a rotation and a translation component [22]. The 3-D surface can also be represented by depth information at zero crossings which have been obtained from a contour detector [24]. Each object can be represented by its 3-D motion, its 3-D structure, and its texture parameters [26], with the object structure representation consisting of a 3-D triangular mesh.

Even though 3-D scene modeling represents an attractive coding approach for stereoscopic sequences, the complexity

of this modeling is often very important. It is also possible that this modeling be inadequate in some image regions because of occlusions, for instance. One advantage of the explicit extraction of depth information is that frontal objects can easily be separated from the background [27]. It then becomes possible to treat these two entities differently, using hybrid approaches combining 2-D and 3-D representations. The selection of an *object* mode (3-D) or a *block* (2-D) mode can be done with the goal of minimizing reconstruction errors [28] or by taking into account the magnitude of depth variations [20].

#### D. The Proposed Compression Approach

Most of the stereoscopic compression methods presented in the literature share a first computational step consisting of local matching between the images of each stereoscopic pair. These methods are therefore confronted with the problems inherent to stereoscopic matching: its combinatorial nature, its sensitivity to variations in illumination conditions and perspective deformations, its sensitivity to the presence of occlusions, and its prohibitive computation time.

To gain wide acceptance in application areas which already make use of monocular imaging, a stereoscopic system should not only be efficient, but also compatible with existing monocular systems so as to reduce upgrade costs. Most of the coding methods presented in the literature are developed for numerical transmission of the stereoscopic information. Consequently, those which attempt to preserve monocular compatibility are developed for the MPEG video standard. Many application domains, however, use analogical transmission. One important example is the domain of telerobotics, where the availability of stereoscopic information could be beneficial. Since a wide range of communication systems are based on the NTSC video standard, it is reasonable to attempt to define stereoscopic information representations that are compatible with the NTSC format. A solution consisting of transmitting the left viewpoint information in the even fields and the one for the right viewpoint in the odd fields is already in use in some commercial systems. However, this approach leads to significant degradations in the vertical and temporal directions, and can certainly be improved by developing compression strategies specific to stereoscopic signals.

Most existing stereoscopic coding methods transmit visual information with the same level of detail over the entire image domain. Nevertheless, the human visual system uses higher spatial and depth resolutions at the center of the visual field and rapidly decreasing resolutions with increasing retinal eccentricity. Based upon this aspect of the human system, it may not be necessary to transmit detailed information over the entire visual field, but only inside a fixation region of limited spatial extent.

In this paper, we present a new compression scheme for stereoscopic sequences whose main characteristics are: 1) the absence of a matching between the stereoscopic channels; 2) the exploitation of a fixation region; and 3) the preservation of spectral compatibility with monocular video signals. The choice of an approach which does not involve a matching between the stereoscopic channels is motivated in part by the problems inherent to stereoscopic matching, but also by the observation that most stereoscopic coding applications are targeted to human observation. It is thus legitimate to leave the matching task to the human visual system rather than invest important computing resources to obtain a result of lower quality. The exploitation of a fixation region does not necessarily require a matching between the two channels. On the other hand, this compression approach involves two fundamental problems: 1) the estimation of the position of the fixation region (i.e., determining where the observer is looking), and 2) the modulation of the coding resolution of different image regions so as to match the spatial and stereoscopic resolutions of the human visual system. The first problem is addressed in Section 11, and the second one in Section 111.

The proposed approach compacts the stereoscopic information into the spectral space of a monocular video channel. This is accomplished by decomposing the stereoscopic information into compact components which are then combined, through spectral manipulation operations. It has been recognized [29], [30] that the spatiotemporal spectrum of an NTSC monocular signal could be used more efficiently, and that NTSC signal quality could be significantly improved by performing appropriate filtering operations on the components before forming the composite video signal. Our approach follows these principles by decomposing, first, the left and right fields of the stereoscopic pairs into low-pass and highpass luminance components. High-frequency components are limited to a fixation region, thus allowing a reduction of their spectral extent. A composite video signal is then formed by positioning the different components into the available spectral space through filtering and modulation. The available spectral space is efficiently used by exploiting the Fukinuki holes [31] to position the chrominance components of both stereoscopic channels into different quadrants of the vertical-temporal frequency plane. The approach is compatible with the NTSC standard in the sense that the same color subcarrier and the same spectral region are used for the chrominance components. It is therefore possible to use standard NTSC decoders to separate the luminance from the chrominance components. Additional processing must then be performed to recover the information corresponding to the left and right viewpoints. This coding approach achieves compression by representing the fixation region more finely than the peripheral area. By doing so, stereoscopic resolution is also implicitly reduced outside this region since the fineness of the correspondence between the left and right channels is reduced there.

Three methods, all based on the above-mentioned principles, are developed to accommodate the bandwidth of different communication systems: 1) the *studio* approach for which the horizontal spectral limit is fixed to half the sampling frequency of the original images, i.e., 7.16 MHz; 2) the *broadcast* approach in which this limit is fixed to 4.2 MHz in accordance with the NTSC standard; and 3) the *integrated* approach which is a flexible representation simultaneously accommodating the bandwidth constraints of the two previous methods. The *studio* format is suitable for use with digital D2 recorders for composite video signals, while the *broadcast* format is

appropriate for image transmission. The three methods yield a 4:1 compression ratio since the composite signal, of the size of a luminance component, combines the two luminance and the four chrominance components (luminance is sampled at  $4f_{sc}$  and chrominance at  $2f_{sc}$ ).

Section II deals with the estimation of the fixation region, and presents a psychophysical study on visual strategies used during the accomplishment of a dynamic depth discrimination task involving stereoscopic perception. Results of this study are taken into account for the development of the spectral compression methods which are described in Section III. Experimental results are presented in Section IV, and a brief discussion concludes the paper.

# 11. FIXATION REGION ESTIMATION

Estimating the location of the fixation point is a difficult problem since high-level attentional processes interact in a complex way with low-level mechanisms controlling the eve movements. The sequence of eye movements highly depends upon the task at hand [32], a fact which further hinders the elaborating of general-purpose prediction strategies. The choice of a particular approach for the estimation of the position of the fixation region therefore highly depends on the application under consideration. For inspection tasks, the fixation region might be fixed and known a priori. In robotics applications, an autonomous system can be equipped with active cameras to track an object of interest so that it always appears in a given image region. For telerobotics applications, eye movement recording systems can be used to directly measure the operator's fixation points [13]. Eye movement tracking devices are also used in 3-D television systems in order to extend the image region over which detailed stereoscopic information is presented [13]. Methods based on eye movement measurements must perform in real time, and often require cumbersome equipment that may weary the user. When such methods cannot be used, it is also possible to predict the fixation region directly from the images content. For instance, in the context of television viewing, a substantial degree of agreement is found between subjects in terms of where they are looking during the viewing of video sequences representative of typical television programs [34]. Another study on visual search reports that different types of fixation sequences are expected, depending on the nature and the complexity of the images over which the search is performed [35].

These studies suggest that characteristic viewing behaviors might be associated with specific visual tasks. However, the few available relevant experimental data about visual strategies pertaining to stereoscopic viewing led us to perform a psychophysical investigation on this topic. Dynamic depth discrimination—the determination of the direction of motion of objects moving in depth—was chosen as the basic visual task for our study since it is typical of applications of stereoscopic viewing to telemanipulation, where a remote effector is continuously repositioned in depth under the control of an operator. A situation where an operator had to accomplish his task in the presence of dynamic distractors was simulated. In order to isolate the stereoscopic modality of the visual system, all stimuli were presented as dynamic random dots stereograms. The main task of the subjects was to indicate whether objects moving in depth in the image center were approaching or receding. Distractors moving in depth, either toward the observer or in a fronto-parallel plane, were following paths end to end of the screen with eight possible directions. As their path crossed the central region, the distractors interfered with the main task, and subjects had the possibility to make them temporarily disappear so as not to mask the main task. Subjects' eye movements were recorded during the task so that visual strategies could be assessed.

The experiment addressed two fundamental aspects of the performance of an operator in a telemanipulation context. The first one concerns the visual strategies used during task execution. It should be expected that, in the absence of distractors, the operator's fixations concentrate in the image region associated with the main task. However, whether or not moving distractors will induce eye movements that bring fixations away from the main region is not clear. If this were the case, a strategy for predicting the fixation region should involve detecting objects in motion, and choosing a location based upon characteristics of the moving objects, such as position, motion type, and direction. On the other hand, if the fixations remain mainly in the central region and distractors are detected but prevented from interfering with the main task, this will indicate that peripheral vision may be playing an active role in the accomplishment of the task. The prediction strategy will, then, simply consist of selecting the largest possible region around the current fixation point in order to maximize the probability that the next fixation point falls into the estimated region.

The second aspect deals with the specific conditions of stereoscopic viewing. When using stereoscopic displays to present visual information in 3-D, it is generally assumed that subjects place their fixation plane on the computer screen [36]. Disparity values are therefore calculated with respect to this reference in order to reconstruct objects with appropriate depth values. However, it is possible that subjects place their convergence plane in front or behind the computer screen. This situation may provide conflictual information to subjects since accommodation and vergence become dissociated. It may also affect the measurement of the fixation points since eye movement recording systems usually also rely on the hypothesis that the fixation plane is positioned on the computer screen, and thus do not evaluate a true fixation position in 3-D space, but only provide a measure of the projection of this point onto the computer screen. This observation suggests that any explanation of experimental data should not neglect the inherent stereoscopic geometry involved in the viewing of the stimuli.

Experimental results show that subjects can be classified into two groups according to the distribution of their fixations. Fig. 1 illustrates the averaged distribution of fixations for the two groups. For the first group [Fig. 1(a)], most of the fixations are concentrated in the region corresponding to the main task, while for the second group [Fig. 1(b)], most of the fixations are in a larger region, containing the central region and extending to the right. Results therefore indicate that subjects can perform the main task and process the dynamic distractors without significantly displacing their fixation point. Distractors are



Fig. 1. Averaged fixation distributions for two groups of subjects. Each figure maps the screen where stimuli images were presented into discrete regions which are represented with intensity values indicating the cumulative number of fixations measured in each region.

probably detected using peripheral vision, which is known to be more sensitive to motion than to spatial details [37].

Our explanation of the difference between the two subjects groups is based upon the geometry involved in stereoscopic viewing. We propose that subjects in the first group were placing their fixation point on the computer screen, while subjects in the second group were sometimes placing their fixation plane in front of the computer screen. When doing so, the fixations measured on the computer screen were horizontally shifted. Our experiment involved measuring left eyes movements, which explains the shift to the right in the fixation distribution.

In summary, two main conclusions can be drawn from this study. First, dynamic depth discrimination tasks can be performed without any significant displacement of the fixation region. Second, in general stereoscopic viewing, subjects place their fixation plane either on the computer screen, as is generally assumed, or in front of it. In the latter case, the fixations on the computer screen are shifted to the right for the left eye, and to the left for the right eye. These conclusions suggest a strategy for the estimation of the fixation region which favors first, the region corresponding directly to the task, and next, a region situated on the opposite side of the viewing eye for each channel. This strategy is incorporated into the stereoscopic compression methods presented in the next section. Because each of these methods must satisfy particular bandwidth constraints and involve fixation regions of different sizes, the application of this strategy is specific to each method, and is thus presented along with the description of the coding methods. On the other hand, for the present study, it was implicitly assumed that a starting fixation point from which fixation displacements could be initiated would be available, either from some *a priori* knowledge of the task or from one of the estimation method reviewed in Section I. The coding methods of the next section also rely on this hypothesis. However, information about the position of the fixation region must be provided for only one channel since an automatic procedure to find the corresponding fixation point in the other channel has been developed. This procedure is common to all coding methods, and is described in the next section.

# 111. STEREOSCOPIC COMPRESSION METHODS

The compression methods presented in this section compact the stereoscopic information into the spectral space of a monocular video channel by exploiting the fact that highfrequency information is not to be transmitted over entire fields, but only within the fixation region. Three compression methods are proposed in order to efficiently exploit the available bandwidth of various communication systems: 1) the studio approach for which the horizontal spectral limit is fixed at half the sampling frequency of the original images; 2) the *broadcast* approach for which this limit is fixed to 4.2 MHz in accordance with the NTSC standard; and 3) the *integrated* approach which is a flexible representation accommodating the bandwidth constraints of the two previous methods. The studio format is suitable, for instance, to be used with digital D2 recorders for composite video signals, while the *broadcast* format is appropriate for image transmission. The three coding methods are described next, following a brief review of the NTSC signal characteristics and a description of the procedure for finding the fixation region position in the second stereoscopic channel.

# A. NTSC Signal Characteristics

The composite NTSC signal has the form

$$U(t) = Y(t) + I(t)\cos(2\pi f_{sc}t + \theta) + Q(t)\sin(2\pi f_{sc} + \theta)$$
(1)

where Y(t)I(t), and Q(t) represent, respectively, the luminance and the two chrominance components, and the color subcarrier  $f_{sc}$  is equal to 3.58 MHz. The phase of  $f_{sc}$  alternates from 0 to  $\pi$  from line to line and from frame to frame. In this paper, we use luminance and chrominance components sampled, respectively, at  $4f_{sc}$  and  $2f_{sc}$ . According to the Nyquist sampling theorem, the spectral space occupied by these components is therefore limited to  $2f_{sc}$  and to  $f_{sc}$ , respectively. However, the narrow NTSC standard specifies a reduced bandwidth of 1.5 MHz for the *I* component and 0.5 MHz for the Q component.

We consider two types of NTSC signals: the *studio* (or *nonbroadcast*) signal, with a luminance component limited to  $2f_{sc} = 7.16$  MHz, and the *broadcast* signal with a luminance component limited to 4.2 MHz. In both cases, it is assumed that the signal contains 480 active lines/frame, which limits the vertical spectral space to  $f_v = 120$  c/ph for each field. Chrominance components occupy a spectral space centered at  $(f_{sc}, f_v)$  in the horizontal–vertical frequency plane, and at (15 Hz,  $f_v$ ) in the temporal–vertical plane. The spectral space available for both types of signals is illustrated in Fig. 2.

The location of the so-called *Fukinuki holes* [31] is illustrated in Fig. 2(c). These holes are not exploited by standard NTSC signals. It is, however, possible to use them to insert additional information by using a second subcarrier whose phase is inverted with respect to  $f_{sc}$  on every alternate field.

#### **B.** Fixation Region Estimation

A fixation region is defined here through its center position, which corresponds to the observer's fixation point, and by its size. The size is limited by the spectral space allocated to the high-frequency information, and thus only the fixation region position must be estimated for each field. In this work, we assume that the location of the fixation region



Fig. 2. Spectral space occupied by one field of an NTSC signal: horizontal-vertical frequency plane for (a) the *studio* case and (b) the *broadcast* case; (c) temporal-vertical frequency plane for both cases. Fukinuki holes are not exploited by standard NTSC signals.

in one of the two channels has been obtained through one of the strategies proposed in the literature and presented in Section 11. The corresponding position in the second channel can be estimated by a one-dimensional (1-D) correlation-based matching procedure applied to an 8 x 8 block A positioned around the fixation region center of the first channel. Block Ais correlated with 8 x 8 blocks  $B_i$  of the second channel, positioned at the same image row, according to (2). The position of the block  $B_i$  which maximizes the correlation measure r provides the best estimate of the fixation region position in the second channel:

$$\Gamma = \frac{\sum_{n_1} \sum_{n_2} A(n_1, n_2) B_i(n_1, n_2)}{\sqrt{\sum_{n_1} \sum_{n_2} A^2(n_1, n_2) \sum_{n_1} \sum_{n_2} B_i^2(n_1, n_2)}}.$$
 (2)

The large size of the fixation region ensures that the actual fixation point remains included in the fixation region, so that positioning errors during the matching process do not significantly affect the overall method. When the estimated position of the fixation region makes some of its pixels fall outside the image boundaries, the fixation region is shifted in order to be aligned with the image. For the validation of the coding methods presented in the next sections, the position of the fixation region is arbitrarily selected in one image of the pair, and estimated in the corresponding channel using the correlation method described above. Contrary to stereoscopic compression algorithms based on disparity compensation, the matching procedure used here is applied only to one block of the entire image.



Fig. 3. Spectral space subdivision for the *studio* method. E'LPL, E'LPR, YHPL, and YHPR represent, respectively, the left and right low-pass and left and right high-pass luminance components.

# C. The Studio Approach: Luminance Component Limited to $2f_{sc} = 7.16 \text{ MHz}$

Two variations of the *studio* method have been investigated. One method requires only spatial filtering; the other necessitates spatiotemporal filtering to exploit the Fukinuki holes [31]. Both methods differ only in the processing of the chrominance components.

I) Luminance Processing: Luminance processing consists of decomposing the luminance component of each channel into a low-pass and a high-pass component, and in combining them to occupy the spectral space as illustrated in Fig. 3. The luminance component of each channel is decomposed into a low-pass and a high-pass component, YLPL and YHPL for the left channel and YLPR and YHPR for the right channel, using 2-D spatial separable FIR filters with cutoff frequencies at  $f_{sc}$  and  $f_v/2$ . The high-pass components YHPL and YHPR occupy much larger spectral spaces than the ones they are allocated in Fig. 3. However, since highfrequency information need not be kept for the entire field, but only for the fixation region, their spectral extent can be reduced by exploiting the scaling property of the Fourier transform. Each high-pass component is thus spatially limited to a region centered at the fixation region position, with the width and height equal to one half of the field's dimensions. However, rather than the estimated fixation region positions, positions shifted to the left for the right viewpoint and to the right for the left viewpoint are used. In accordance with the conclusions of our analysis of the fixation process, this shift is performed in order to locate a larger portion of the fixation region on the opposite side of the viewing eye, and to augment the probability that the observer actual fixation point will fall into the fixation region (see Section 11). The amount of shift that is performed is also based on our fixation process study, and corresponds to  $3.5^{\circ}$  of visual angle for a standard viewing distance of 52 cm. The high-pass components YHPL and YHPR are then spatially enlarged (zoomed) to the original field size by expanding them by a factor of two in the horizontal and vertical directions, and filtered at  $f_{sc}$  and  $f_v/2$  to eliminate spectral repetitions. Finally, the low-pass and high-pass components are modulated and appropriately filtered to remove spectral duplications introduced by the modulation operation, so as to position them into spectral space as illustrated in Fig. 3.



Fig. 4. Spectral positioning of the chrominance components according to the *studio* method: (a) spatial positioning and (b)–(c) spatiotemporal positioning. When using the spectral Fukinuki holes, the chrominance components of the left and right channels are positioned in the same spectral region in the horizontal–vertical frequency plane (b), but are separated in the temporal–vertical frequency plane (c).

2) Chrominance Processing: The available spectral space for the chrominance component in Fig. 3 extends horizontally from  $f_{sc} - f_{sc}/2$  to  $f_{sc} + f_{sc}/2$  and vertically from  $f_v - f_v/4$  to  $f_v$ . With the narrow NTSC standard, the I and Q components, limited, respectively, to 1.5 and 0.5 MHz, are amplitude modulated in quadrature with vestigial sideband such that their spectral extent does not exceed 4.2 MHz. With the studio approach, some extra spectral space is available for the chrominance components between 4.2 and 5.37 MHz, but four chrominance components must be fitted into the overall chrominance space. The first proposed method consists of modulating in quadrature with vestigial sideband the Icomponents of the two channels to occupy the spectral space between 1.79 and 4.2 MHz, and of modulating in quadrature the two Q components to occupy the space between 4.2 and 5.37 MHz. The spectral positioning of the chrominance components according to this method is illustrated in Fig. 4(a).

Improvement on this first method is possible by exploiting more efficiently the available spectral space in the vertical-temporal frequency plane. The regions referred to as the *Fukinuki holes* [31] in the first and third quadrants of Fig. 2(c) do not contain any chrominance information in the NTSC standard. The second proposed approach consists of using these holes for the chrominance information of one of the two stereoscopic channels. The NTSC subcarrier  $f_{sc}$  is used to amplitude modulate in quadrature the I and Q components of one channel, and a second subcarrier  $f'_{sc}$ , whose phase is inverted with respect to  $f_{sc}$  in every alternate field, amplitude modulates in quadrature the I and Q components of the other channel. This approach implies that temporal filtering must be performed at the receiver to recover the four chrominance components. The spectral positioning of the chrominance according to this approach is illustrated in Fig. 4(b) and (c).



Fig. 5. Spectral space subdivision for the *broadcast* method. YLPL, YLPR, YHP1L, YHP1R, YHP2L, and YHP2R represent, respectively, the left and right low-pass and the high-pass luminance components of both channels.

# D. The Broadcast Approach: Luminance Component Limited to 4.2 MHz

The compression strategy in the *broadcast* approach basically follows the same principles as for the *studio* case, except that the spectral space is now limited horizontally to 4.2 MHz. Consequently, the horizontal size of the fixation region is limited to one fifth of the field size, which is not large enough to benefit from the shifting of the fixation region suggested by the psychophysical findings of Section 11. Therefore, the estimated fixation position is directly used.

1) Luminance Processing: The luminance component are decomposed into a low-pass and a high-pass component, YLPL and YHPL for the left channel and YLPR and YHPR for the right channel, using 2-D spatial separable FIR filters with cutoff frequencies at 2.1 MHz and  $f_v/2$ . The high-pass components are separated into two subcomponents: the vertical high frequencies with horizontal frequencies less than 2.1 MHz, YHP1L and YHP1R, and the horizontal high frequencies above 2.1 MHz, YHP1L and YHP1R, and the horizontal high frequencies are spatially limited to the fixation region, and then enlarged by a horizontal factor of five and a vertical factor of two. The low-pass and high-pass components of both channels are then combined to occupy the spectral space as illustrated in Fig. 5.

2) *Chrominance Processing:* The processing of the chrominance components can be performed in two different ways. The first one consists of exploiting the Fukinuki holes to transmit the chrominance components of the two channels with vestigial sideband, and implies temporal filtering at the receiver to separate the chrominance components of the two channels. The spectral positioning of the chrominance components according to this method is done in a manner similar to the *studio* approach (see Fig. 4).

The second approach is appropriate when temporal filtering cannot be used. The bandwidth of the *I* component is reduced to 0.5 MHz, and the chrominance components of one channel are amplitude modulated in quadrature to occupy the spectral space which extends horizontally from 3.08 to 4.08 MHz. Another subcarrier f'' = 2.58 MHz is used to amplitude modulate in quadrature the chrominance components of the second channel, so that they occupy the spectral space which extends horizontally from 2.08 to 3.08 MHz.



Fig. 6. Spectral space subdivision for the *integrated* method. YLPL, YLPR, YHP1L, YHP1R, YHP2L, YHP2R, YHP3L, YHP3R, YHP4L, and YHP4R represent, respectively, the left and right low-pass and the high-pass luminance components of both channels.

# E. The Integrated Approach: Luminance Component Limited to 4.2 or 7.16 MHz

It would certainly be useful to have a generic signal representation that is simultaneously suitable for the *studio* and the *broadcast* approaches. Whenever the system bandwidth is large enough to accommodate *studio* signals, a maximal amount of stereoscopic information should be conveyed by the composite video signal. Whenever the bandwidth can only accommodate *broadcast* signals, the same video signal filtered at 4.2 MHz should convey the essential stereoscopic information. The compression scheme for the *integrated* approach follows this idea by extending the *broadcast* approach to incorporate a more sophisticated processing of the fixation region which takes into account the psychophysical findings reported in Section 11.

Given the fixation region position in each stereoscopic channel, this region is divided into two subregions: a central and an adjacent subregion, with the central subregion position corresponding to the fixation region position. For the left viewpoint, the adjacent subregion is located to the right of the central subregion, and vice versa for the right viewpoint. The size of each subregion corresponds to one fifth of the horizontal original field size, for an overall fixation region size of two fifths of the original field size. This division of the fixation region into two subregions takes into account the psychophysical findings of Section II since a larger portion of the overall fixation region (the adjacent subregion) is on the opposite side of the viewing eye for each channel.

The partitioning of the spectral space according to the *integrated* approach is illustrated in Fig. 6. The spectral occupation from 0 to 4.2 MHz is the same as for the *broadcast* approach with Fukinuki holes, with the fixation region of the *broadcast* approach corresponding to the central subregion of the *integrated* approach. The spectral space between 4.2 and 7.16 MHz is used to insert additional high-frequency information corresponding to the adjacent subregions of the fixation region, to extend the chrominance space to 5.37 MHz, and to improve the quality of the right low-pass component. The adjacent subregion for each channel is processed in a way similar to the central subregion: the high-pass component is subdivided into two subcomponents representing the vertical

high frequencies above  $f_v/2$  and below 2.1 MHz, YHP3Land YHP3R, and the horizontal high frequencies above 2.1 MHz, YHP4L and YHP4R. These subcomponents are spatially limited to the adjacent subregion, enlarged, filtered, and modulated to occupy the available spectral space as illustrated in Fig. 6. The right low-pass component is now modulated with double sideband so that reconstruction quality is improved when the full bandwidth is available. The chrominance components are amplitude modulated in quadrature with Fukinuki holes (see Fig. 4). When the bandwidth is limited to 4.2 MHz, the right low-pass component and the chrominance components become modulated with vestigial sideband, and the high-frequency information corresponding to the adjacent subregions is not transmitted, as for the *broadcast* case.

## F. Summary of the Three Methods

The *integrated* approach represents a compromise between the *studio* and the *broadcast* approach. For a bandwidth of 7.16 MHz, the fixation region width with the *integrated* approach is equal to 2W/5, where W represents the width of one field, and the bandwidth of the low-pass components corresponds to 2.1 MHz, while with the *studio* method, these two parameters are equal to W/2 and 3.58 MHz. This reduction in fixation region size and bandwidth is, however, compensated by an increased flexibility which makes an *integrated* video signal suitable for a bandwidth of 4.2 MHz by low-pass filtering it.

The component representation introduced in this paper represents an efficient way to compact the information since the essential high-frequency information (corresponding to the fixation region) occupies a reduced spectral space. Components combination by means of spectral manipulations to form a composite video signal suitable for analogical transmission represents one possible way to take advantage of the this representation. However, it should be clear that the the proposed representation is not limited to an analogical transmission of the information, but can also be suitable for its numerical transmission. Future work on this topic could use the proposed component representation as a preprocessing step before feeding the components, represented either individually or as a composite signal, into standard numerical encoders, such as JPEG or MPEG encoders. Two possible ways to exploit the proposed component representation for numerical transmission are proposed. First, since the high-frequency information has been manipulated to be represented in a bandlimited spectral space, it is expected that the left and right luminance and chrominance components of a field will all be represented independently with compact codes. Second, using a numerical encoder for the composite video signal should further increase the 4:1 compression introduced by this representation.

# IV. RESULTS

Five stereoscopic sequences containing several types of motion and depth plane variations have been compressed and reconstructed according to the various methods described in Section 111. For the *aqua* sequence, the camera executes a slow traveling in front of an aquarium containing static hung plastic



Fig. 7. Reconstructed luminance components according to the *studio* method. (a) Reconstructed low-pass component. (b) Reconstructed high-pass component. (c) Reconstructed enlarged fixation region (for images (b)–(c), contrast enhancement techniques have been applied to facilitate visualization).

fishes located at different depth planes. The fixation region was chosen to be on the big fish located on the bottom left of the aquarium because of its color details. The roundabout sequence represents a rotating roundabout for children. Various movements causing some occlusions are present, and due to the rotation, the toys on the roundabout have a depth motion component. The fixation region was chosen to be one of the toys of the roundabout, the sailboat, because it undergoes a large displacement during the sequence. The *tunnel* sequence contains a toy train which enters a tunnel following a curved track, and thus the receding motion causes some deformations of the wagons apparent shape. The fixation region was chosen to be on the personage in front of the horse near the tunnel since it seems that, during the sequence acquisition, the cameras were focused on this region. The piano sequence contains a man playing the piano with his fingers and his shirt moving in a nonrigid way. His highly textured shirt was chosen as the fixation region because of the fine details it contains. The *train* sequence contains two trains going in opposite direction with different speeds which pass each other. Some parts of one of the train are thus temporarily occluded. The fixation region was chosen to be on the barrels on one of the trains since they are the most detailed objects on the trains that are visible throughout the entire sequence.

All spatial filtering operations were performed using separable FIR filters with a number of coefficients ranging from 11 to 41. Temporal filtering for the exploitation of the Fukinuki holes was performed over three fields. For illustrative purposes, intermediate results obtained during the treatment of one stereoscopic sequence with the *studio* method are presented. Fig. 7(a) and (b), respectively, illustrates the reconstructed low-pass and high-pass luminance components of the two first left fields of the *tunnel* sequence, with the region of fixation centered on personage in front of the horse (center left). The enlarged high-pass component corresponding to the area of fixation is shown in Fig. 7(c). This component was scaled down to form Fig. 7(b).

The proposed compression approaches have been compared using PSNR measures. Results are tabulated for the second image of each sequence. The first two columns of Table I present PSNR values calculated between original luminance and reconstructed low-pass luminance components over the entire images. These measures indicate the quality level of the reconstructed luminance signal in the peripheral region. PSNR values are higher for the *studio* approach since the horizontal cutoff frequency for the low-pass luminance components is higher than with the other methods.

The last two columns of Table I present PSNR values between the original luminance and the reconstructed luminance in the fixation region. PSNR values range from 27 to 35 dB for the *studio* method and between 23.5 and 33 dB for the other methods. The better performance of the *studio* method can be explained by the fact that fewer filtering and modulation/demodulation operations are involved. The obtained PSNR values have been compared with PSNR values for the reconstructed low-pass components in the fixation region (not the ones reported in Table I which are for the entire images). This comparison shows an improvement in signal quality of 2.5–5.9 dB when the high-frequency information is added to the low-pass components. The low PSNR value

 TABLE I

 PSNR Between original Luminance and Reconstructed Low-Pass

 Luminance components (First 2 columns) and Between Original

 Luminance and Reconstructed Low-Pass + High-Pass Luminance

 Components in the Fixation Region (Last 2 c o l u m n s )

	YLPL	YLPR	Y L	Y R
Tunnel stu.	27.35	27.31	32.01	32.34
Dinno etu	26.87	29.24	27.12	29.17
<b>A</b>	28.65	26.75	33.95	32.05
Round. stu.	27.24	28.94	31.71	31.89
Train stu.	31.48	31.40	34.62	34.30
Tunnel broad.	26.67	26.43	29.91	29.93
Piano broad.	25.04	27.14	23.69	25.83
Aqua broad.	26.80	25.82	29.72	27-72
Round. broad.	25.56	26.95	28.13	29.44
Train broad.	31.00	30.64	31.60	31.20
Tunnel int. (7.16)	26.70	26.51	30.87	30.47
Piano int. (7.16)	25.08	27.26	23.95	-27-61
Aqua int. (7.16)	26.87	24.94	29.77	27.30
Round. int. (7.16)	25.60	27.09	29.36	29.53
Train int. (7.16)	31.03	30.63	33.01	32.28
Tunnel int. (4.2)	26.70	26.41	29.93	29.38
Piano int. (4.2)	25.07	27.01	23.70	25.09
Aqua int. (4.2)	26.87	24.80	29.86	27.16
Round. int. (4.2)	25.59	26.85	28.14	28.98
Train int. (4.2)	31.03	30.59	31.66	30.82
	•			•

obtained for the *piano* sequence with the *broadcast* approach and the *integrated* approach limited to 4.2 MHz is explained by the fact that the low-pass components in the fixation region used with these methods have low PSNR values: 20.43 and 23.33 dB for the original low-pass components, and 19.80 and 22.15 dB for the reconstructed low-pass components. Reconstructed luminance with PSNR values of 23.69 and 25.83 dB therefore represent improvements in signal quality of 3.89 and 3.68 dB, which is comparable to what is observed with the other sequences.

PSNR measures have also been computed between original and reconstructed chrominance components. PSNR values for the reconstructed I component range between 33.5 and 36 dB when its bandwidth is limited to 1.5 MHz and between 32 and 34 dB when it is limited to 0.5 MHz. PSNR values for the Q component range between 34 and 38.5 dB. Modulation with Fukinuki holes with the vestigial or double-sideband produce only small differences in PSNR values of the reconstructed components, but both methods are better (by more than 1 dB on average) than modulation without Fukinuki holes.

Results show that all of the proposed coding approaches lead to an improvement in reconstructed signal quality in the fixation region, as desired. The *studio* approach is simpler, leads to better signal quality reconstruction than the other approaches in the fixation and peripheral regions, but also requires a larger bandwidth to transmit the stereoscopic information. The *integrated* approach is a flexible representation that allows the -transmission of the stereoscopic information with a 7.16- or a 4.2-MHz communication channel. However, compared to the studio approach, the fixation region size is reduced, and the signal quality in both the fixation and the peripheral region is diminished. A transmission of the chrominance components exploiting the Fukinuki holes should be favored. However, when temporal filtering is not possible at the receiver, the modulation scheme for the chrominance components proposed with the studio approach also turns out to be adequate.

#### V. CONCLUSION

A new compression scheme for interlaced stereoscopic sequences has been presented. It exploits the presence of a fixation region to compact the stereoscopic information into the spectral space of monocular video channel. Two fundamental problems associated with the proposed coding approach were addressed: 1) the determination of the position of the fixation region, and 2) the modulation of the level of details preserved in the images as a function of the fixation region.

Possible approaches for the estimation of the fixation region were reviewed, and a psychophysical study on visual strategies used during the accomplishment of a dynamic depth discrimination task was presented. Results of this study suggest that, in order to maximize the probability that the observer actual fixation points falls into the predicted fixation region, this region should be shifted with respect to its estimated image position so that a larger portion of the region is located on the opposite side of the viewing eye for each stereoscopic channel. Three compression methods which take into account these psychophysical results have been proposed to fully exploit the available communication bandwidth: the *studio* approach in which the horizontal spectral space is limited to half the sampling frequency of the images, the broadcast approach in which the horizontal spectral limit is fixed at 4.2 MHz, in accordance with the NTSC standard, and the *integrated* approach which is simultaneously suitable for both conditions.

The basic principles of the proposed coding methods consist of 1) decomposing the left and right fields of the stereoscopic pairs into low-pass and high-pass luminance components; 2) limiting high-frequency components to a fixation region to reduce their spectral extent; and 3) forming a composite video signal by positioning the different components into the available spectral space through filtering and modulation. The processing of the chrominance components can be performed in the spatial domain or in the spatiotemporal domain if the Fukinuki holes are exploited. The approach is compatible with the NTSC standard in the sense that the same color subcarrier and the same spectral region is used for the chrominance components.

Experimental results to validate the proposed approach were presented. All filtering operations were implemented using separable FIR filters with a relatively small number of coefficients. The methods are therefore suitable for an efficient hardware implementation. Future works include the development of fixation region prediction strategies for generic visual tasks, as well as an optimization of the filters in order to reduce the number of coefficients and improve the quality of the reconstructed images.

#### REFERENCES

- A. Liu, G. Tharp, and L. Stark. "Depth cue interaction in telepresence and simulated telemanipulation." in *Proc. SPIE: Human Vision, Visual Processing, Digital Displ. III*, vol. 1666, 1992. pp. 541–547.
- [2] M. Hirose, K. Hirota, and R. Kijima, "Human behavior in virtual environments." in *Proc. SPIE: Human Vision, Visual Processing, Digital Displ. III*, vol. 1666, 1992, pp. 548–559.
- [3] Y.-Y. Yeh and L. D. Silverstein. "Spatial judgements with monoscopic and stereoscopic presentation of perspective displays," *Human Fact.*, vol. 34, pp. 583–600, Nov. 1992.
- [4] A. K. Jain. Fundamentals of Digital Image Processing. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [5] H. G. Musmann. P. Pirsh. and H.-J. Grallert. "Advances in picture coding." *Proc. IEEE*. vol. 73. pp. 523–548, Apr. 1985.
- [6] T. Ozkan and E. Salari, "Coding of stereoscopic images," in *Proc. SPIE: Image and Video Processing*, San Jose. CA. vol. 1903. Feb. 1993, pp. 228–235.
- [7] I. Dinstein, G. Guy, J. Rabany, J. Tzelgov. and A. Henik. "On stereo image coding," in *Proc. Int. Conf. Pattern Recognition*, Rome, Italy. vol. 1, Nov. 1988, pp. 357–359.
- [8] J. Steurer, "Coding stereoscopic images for the human observer." in Proc. European Workshop 3D-TV Signal Processing. Rennes, France. Nov. 1992.
- [9] H. Yamaguchi, Y. Tatehira. K. Akiyama, and Y. Kobayashi, "Stereoscopic images disparity fo predictive coding," in *Proc. Int. Conf. Acoust.*, *Speech, Signal Processing*, Glassgow, Scotland. vol. 3, May 1989, pp. 1976–1979.
- [10] A. Schertz. "Source coding of stereoscopic television pictures," in *Proc. IEE Int. Conf. Image Processing Appl.*, Maastrict, The Netherlands, Apr. 1992, pp. 462-464.
- [11] A. Arditi, "Binocular vision," in *Handbook Perception and Performance: Sensory Processing and Perception*. K. R. Boff, L. Kaufman. and J. P. Thomas. Eds., vol. 1. New York: Wiley. 1986, pp. 23.1–23.40.
  [12] L. A. Olzak and J. P. Thomas, "Seeing spatial patterns." in *Handbook*
- [12] L. A. Olzak and J. P. Thomas, "Seeing spatial patterns." in *Handbook Perception and Performance: Sensory Processing and Perception*, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds., vol. 1. New York: Wiley, 1986, pp. 7.1–7.53.
- [13] P. Obrador. "Eye-tracked based stereo video compression scheme," in Proc. Int. Workshop Stereoscopic and 30 Imaging, Santorini. Greece, Sept. 1995, pp. 64–69.
- [14] M. G. Perkins, "Data compression of stereopairs." *IEEE Trans. Commun.*, vol. 40, pp. 684-696, Apr. 1992.
  [15] Y.M. Zhu, Y. Gao, and R. Goutte. "Stereo image compression by the
- [15] Y.M. Zhu, Y. Gao, and R. Goutte. "Stereo image compression by the combined use of the {DCT} and binocular fusion and rivalry." in *Proc. Int. Festival 30 Imaging*, Paris, France, Sept. 1991. pp. 203–211.
  [16] I. Dinstein, M. G. Kim, and A. Henik, "Compression of stereo images
- [10] I. Dinstein, M. G. Kim, and A. Henik, "Compression of stereo images using subsampling and transform coding," *Opt. Eng.*, vol. 30. pp. 1359–1364, Sept. 1991.
- [17] L. Oisel, L. Morin, and C. Labit, "Stereoscopic data compression using projective geometry," in *Proc. Int. Workshop Stereoscopic and* 3D Imaging, Santorini, Greece. Sept. 1995, pp. 46–51.
- [18] D. De Vleeschauwer. "Coding of the disparity field and the correction signal in predictive 3DTV coding," in *Proc. Int. Festival 3D Imaging*, Paris. France, Sept. 1991, pp. 213–221.
- [19] F. Chassaing, B. Choquet, and D. Pelé, "A stereoscopic television system (3D-TV) and compatible transmission on a MAC channel (3D-MAC)," Signal Processing: Image Commin. 'vol. 4, pp. 3343, Nov. 1991.
- [20] A. Kopernik and D. Pelé, "Multi-mode object-based 3D sequence analysis," in *Proc. European Workshop 3 0 TV*, Rome. Italy. Oct. 1993, pp. 143-147.

- [21] M. Ziegler, "Region-Based Analysis and Coding of Stereoscopic Video." Ph.D. dissertation, Tech. Univ. of Delft, Oct. 1997.
- [22] J. L. Dugelay and D. Pelé, "Motion and disparity analysis of a stereoscopic sequence: Application to 3DTV coding," in *Proc. European Signal Processing Conf.*, Brussels, Belgium, vol. 3, Aug. 1992, pp. 1295-1298.
- [23] S. Malassiotis and M. G. Strintzis, "Object-based coding of stereo image sequences using 3D models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 892–906, Dec. 1997.
- [24] H. Morikawa and H. Harashima. "3D structure extraction coding of image sequences," J. Vision Commun. and Image Representation. vol. 2, pp. 332-344, Dec. 1991.
- [25] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-based coding of stereo image sequences using joint 3-D motion/disparity estimation." *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 312–328, Apr. 1997.
- [26] L. Falkenhagen, "3-D object-based depth estimation from stereoscopic image sequence," in *Proc. Int. Workshop Stereoscopic and 3D Imaging*, Santorini. Greece. Sept. 1995, pp. 81–86.
- [27] M. Waldowski, "A new segmentation algorithm for videophone applications based on stereo image pairs," *IEEE Trans. Commun.*, vol. 39, pp. 1856–1868, Dec. 1991.
- [28] N. Grammalidis, S. Malassiotis, D. Tzovaras, and M. G. Strintzis. "3-D motion estimation and compensation for stereoscopic image sequence coding," in *Proc. European Workshop 3D TV*, Rome, Italy, Oct. 1993. pp. 191–197.
- [29] M. A. Isnardi, "Exploring and exploiting subchannels in the NTSC spectrum," SMPTE J., vol. 97, no. 7, pp. 526–532, 1988.
- [30] É. Dubois and W. F. Schreiber, "Improvements to NTSC by multidimensional filtering," SMPTE J., vol. 97, pp. 446–463, June 1988.
- [31] T. Fukinuki and Y. Hirano, "Extended definition TV fully compatible with existing standards," *IEEE Trans. Commun.*, vol. 32, pp. 948–953, Aug. 1984.
- [32] A. Yarbus, Eye Movements and Vision. New York: Plenum, 1967.
- [33] N. Tetsutani. F. Kishino, and Y. Nagashima, "Stereoscopic display method employing eye-position tracking and HDTV LCD-projector." in *Proc. Int. Workshop HDTV'92*, Kawasaki, Japan, vol. 2, Nov. 1992.
- [34] L. B. Stelmach, W. J. Tam, and P. J. Hearty, "Static and dynamic spatial resolution, in image coding: An investigation of eye movements," in *Proc. SPIE: Human Vision, Visual Processing, and Digital Displ. II*. San Jose, CA, vol. 1453, Feb. 1991, pp. 147–152.
- [35] L. W. Stark, K. Ezumi, T. Nguyen, R. Paul. G. Tharp, and H. I. Yamashita, "Visual search in virtual environments." in *Proc. SPIE: Human Vision, Visual Processing, and Digital Displ. III*, 1992. vol. 1666. pp. 577-589.
- [36] L. F. Hodges and E. T. Davis, "Geometric considerations for stereoscopic virtual environments." *Presence*, vol. 2, no. 1, pp. 34-43, 1993.
- [37] S. Anstis, "Motion perception in the frontal plane: Sensory aspects." in Handbook Perception and Performance: Sens. Processing and Perception, K. R. Boff. L. Kaufman, and J. P. Thomas, Eds., vol. 1. New York: Wiley, 1986. pp. 16.1–16.27.