WATERMARKING FOR LIGHT FIELD RENDERING[1]

Alper Koz, Cevahir Çığla and A. Aydın Alatan

Department of Electrical and Electronics Engineering, METU
Balgat, 06531, Ankara, TURKEY.
e-mail: koz@metu.edu.tr, cevahir@eee.metu.edu.tr, alatan@eee.metu.edu.tr

ABSTRACT

The recent advances in Image Based Rendering (IBR) have pioneered freely determining the viewing position and angle in a scene from multi-view video. Remembering that a person could also record a personal video for this arbitrarily selected view and misuse this content, it is apparent that copyright and copy protection problems also exist and should be solved for IBR applications, as well. In our recent work [1], we propose a watermarking method, which embeds the watermark pattern into every frame of multi-view video and extracts this watermark from a rendered image, generated by the nearest-interpolation based light-field rendering (LFR) and watermark detection is achieved for the cases in which the virtual camera could be arbitrarily located on the camera plane only. This paper presents an extension to the previous formulation for the rendered images where the location of the virtual camera could be completely arbitrary in this new formulation. The results show that the watermark could be extracted successfully for LFR via nearest neighborhood interpolation for any imagery camera location and rotation, as long as the visual quality of the rendered image is preserved.

Index Terms— Watermarking, Free View Television, Light Field Rendering, Image Based Rendering

1. INTRODUCTION

Image-based rendering (IBR) has been developed in the last ten years, as an alternative to the traditional 3-D geometry-based rendering techniques. IBR aims to produce a projection of a 3-D scene corresponding to an arbitrary view-point by using a number of original camera views of the scene. IBR has shown to yield more natural views, compared to the traditional 3-D geometry-based methods [2]-[3]. Due to its advantages, IBR has attracted much attention from the research community, as well as the industry, even leading to some new consumer products, *free-view TV,* where the viewing position and angle can be selected as a result of the application of IBR to the transmitted multi-view video [4]-[5] .

Noting that a viewer might also record a personal video from the arbitrarily selected views and misuse this content, it is apparent that copyright and copy protection problems should also exist and be solved in any IBR application. Among many alternative digital rights management methods, the copyright problem for the visual data has already been approached by embedding hidden imperceptible information, called *watermark,* into the image and video content [6]-[8]. Hence, *watermarking* can be assumed as a good candidate for the solution of the copyright problem for such applications, as well. In this new application area, the watermark should not only be resistant to image-based rendering operations, but it should also be extracted from a generated video for an arbitrary view, which requires an estimation procedure for the imagery camera position and orientation of the rendered view.

In the literature, the most well known and preferred IBR representation is the *light field* approach [9], due to its simplicity and natural-looking outputs. Light field-based methods only require the original images to construct the imagery views [5]. It should be noted that the proposed watermarking method is specially tailored for the systems, whose arbitrary views are generated by using *light field rendering* (LFR).

In *LFR*, one of the main interpolation techniques is *nearest neighborhood interpolation* where the intensity value of each pixel for the virtual view is generated from the pixel intensities belonging only to the nearest camera. This interpolation technique can be specially preferred for rendering due to its simplicity in the implementation. In our recent pioneering work [1], a solution to extract the watermark from the rendered image where the virtual camera could be arbitrarily located on the camera plane only. This paper presents an extension where the location of the virtual camera could be completely arbitrary.

2. PROPOSED WATERMARKING METHOD

In *LFR*, a light ray is indexed as (u_o,v_o,s_o,t_o), where (u_o,v_o) and (s_o,t_o) are the intersections of the light ray with the two parallel planes namely, *camera* (*uv*) plane and *focal* (*st*) plane. These planes are usually discretized, so that a finite number of light rays could be recorded. If all the discretized points from the focal plane are connected to one point on the camera plane, an image (2D array of light fields) is resulted [9]. 4D representation of the light field can also be interpreted as a 2D image array, as it is shown in Fig. 1. In the proposed method, the same watermark is embedded to each image in this 2D image array.

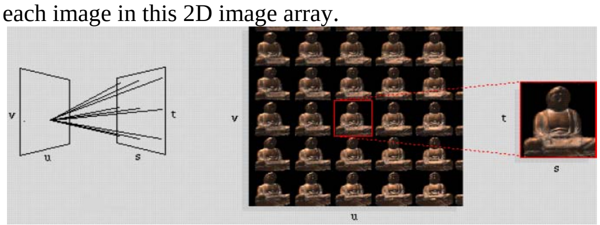

Fig. 1 A sample light field image array: *Buddha* light field [9]

[1] This work is supported by EC within FP6 under Grant 511568 with the acronym 3DTV.

2.1. Watermark Embedding

The proposed method embeds the watermark into each image of the light-field slab by exploiting spatial sensitivity of HVS [1]. For that purpose, the watermark is modulated with the resulting output image, after filtering each light field image by a 3x3 high pass filter and spatially added to that image. This operation decreases the watermark strength in the flat regions of the image, in which HVS is more sensitive, whereas increases the embedded watermark energy in the detailed regions, where HVS is insensitive.

There are two critical points at the embedding stage. First of all, the watermark is embedded to the light-field images, which are the sheared perspective projection of the original camera frames. These frames can be easily obtained by the camera calibration information. Secondly, the watermark component, added to the intensity of each image pixel, is determined according to the intersection of the light ray corresponding to that pixel with the focal plane. The same watermark component is added to the pixels of different camera frames, whose corresponding light rays are intersected at the same point in the focal plane, as illustrated in Fig. 2. The rationale behind such a procedure is to avoid facing with the superposition of the different watermark samples from different camera frames in the interpolation operations during the rendering.

The method applies the following relation to each light field image:

$$I_{uv}^{*}(s,t) = I_{uv}(s,t) + \alpha . H_{uv}(s,t) . W(s,t) \qquad (1)$$

where I_{uv} is the light field image corresponding to the camera at the *(u,v)* position on the camera plane, H_{uv} is the output image after high pass filtering, α is the global scaling factor to adjust the watermark strength, *W* is the watermark sequence generated from a Gaussian distribution with zero mean and unit variance, and finally, I_{uv}^{*} is the watermarked light field image.

2.2 Watermark Detection

The well-known correlation-based detection scheme could be utilized for watermark extraction. Assuming that the position and rotation for the imagery view are known a *priori* (not available in practice), the first step is applying the same rendering operations during the generation of an arbitrary view to the watermark pattern, *W*, in order to generate a "rendered watermark", W_{ren}. After the arbitrary view is filtered by a high pass filter, the normalized correlation between the resulting image and rendered watermark is determined. In the next step, normalized correlation is compared to a threshold for the detection of the watermark. Overall structure of the watermark detection is shown in Fig. 3.

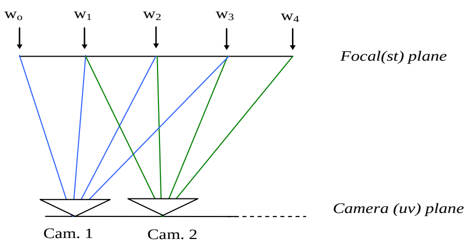

Fig. 2 Watermark embedding methodology

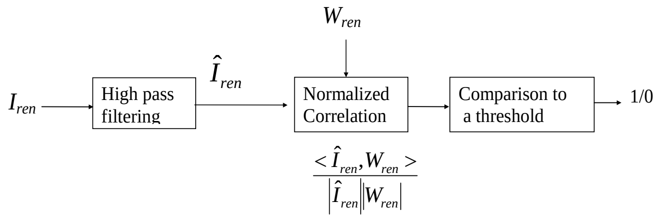

Fig. 3 Overall structure of Watermark Detection Process

2.3 Robustness Results

In order to observe the robustness of watermark detection against IBR effects for the known-view scenario, simulations are performed on *Buddha* [10] and *Teapot* light fields [11]. For the case, in which the arbitrary view is obtained, corresponding to a camera, whose image plane is parallel to camera and focal planes, the original and watermarked rendered views are presented in Fig. 4 for *Buddha* Light field. There is no visible difference between the original and rendered views. The normalized correlation values for the embedded watermark and other 100 randomly generated watermarks from a zero mean and unit variance Gaussian distribution are also shown in Fig. 4.c. The higher value of true watermark correlation with respect to the other values clearly indicates that the watermark can be detected. The correlation results for the other virtual camera positions and orientations are given in [1].

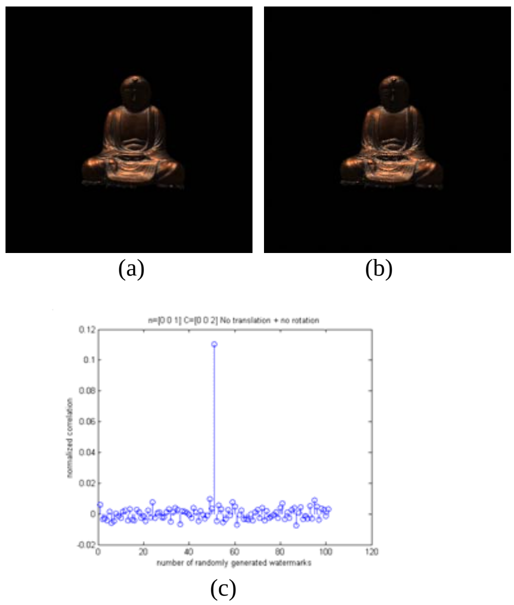

Fig. 4 a) Rendered view b) Watermarked view c) Correlation result (PSNR between rendered and watermarked view: 41.57 dB)

3. AN ANALYSIS ON THE RELATION BETWEEN THE EMBEDDED WATERMARK& RENDERED WATERMARK

The small distortions in the imagery camera position and rotation significantly degrade the detection performance. In order to handle this problem, the relation between the embedded watermark and the rendered watermark is analyzed in three cases:

1. The imagery camera is located in the camera plane and its rotation matrix is a unit matrix. The configuration is shown in Fig. 5. The same watermark, *W*, is embedded to the off-sheared

images of Cameras 1 and 2. From the analysis of the operations in light field rendering, the watermark corresponding to the imagery camera, W_{ren}, will be a *shifted version* of W.

2. The imagery camera is again located in the camera plane and its rotation is not unity. (see Fig. 5.). For this case, the relation between W_{ren} and W is a planar projective transformation [9].
3. The imagery camera is in an arbitrary position and its rotation is not unity (Fig. 5). The relation between W_{ren} and W is again planar projective transformation. However, there will be a scaling between the watermarks which makes the watermark detection more difficult. A possible shift can also occur in this case. This is the general case for the imagery camera position and rotation.

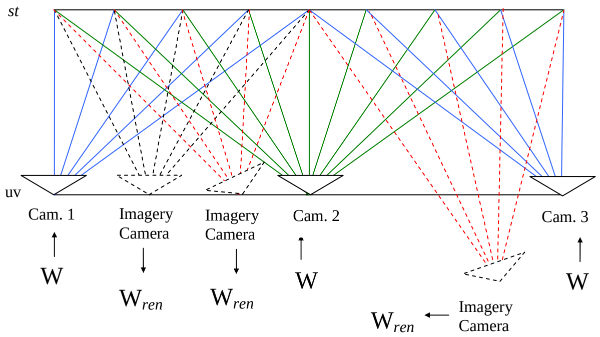

Fig. 5 Configurations for the imagery camera position and rotation from left to right for Case 1, 2 and 3, respectively.

3.1 Proposed Solution for Case I.

In order to solve the problem for the first case, the correlation is computed for the possible shifts of the watermark pattern in the detector [1]. The computation of this operation is decreased by utilizing a Fast Fourier Transform (FFT) and its inverse (IDFT). Specifically, symmetrical phase only matched filtering (SPOMF) is used for correlation computations [8]. The experimental results are presented in [1]

3.2 Proposed Solution for Case II

In Case II, the rendered image is generated from the image corresponding to the nearest neighbor camera. The relation between the rendered and the nearest neighbor original image can be approximated as a planar projective transformation [9]. This property can be used to determine the original image from which the rendered image is generated. The proposed solution and the experimental results are given in [1].

3.3 Proposed Solution for Case III

In this general case, the rendered image is composed of different regions where the light fields for each region are generated from a different light field image corresponding to a different camera. An illustration is given in Fig. 6. The rendered image is formed of three regions and the light fields (pixels) for the first, second and third region are generated from the light fields (pixels) of first, second and third camera, respectively. Similar to the second case, the relation between each region of the rendered image and the corresponding light field image can also be approximated as a planar projective transformation (homography).

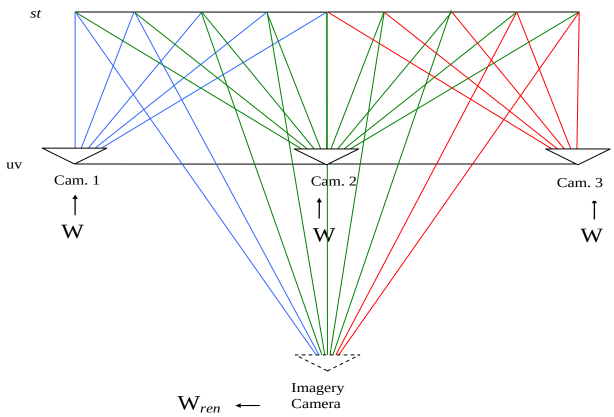

Fig. 6 Each region in different color is generated from a different light field image corresponding to a different camera

The rendered image generated for an imagery camera located at [0 0 2.4] with [0 0 1] orientation is shown in Fig. 7 for the *Buddha* light field [11]. This rendered image is composed of four main regions each of which is generated from the light field images, *buddha.15.15*, *buddha.16.15*, *buddha.15.16* and *buddha.16.16*. The relation between each region of rendered image and the corresponding regions in light field images is a planar projective transformation (Fig. 7).

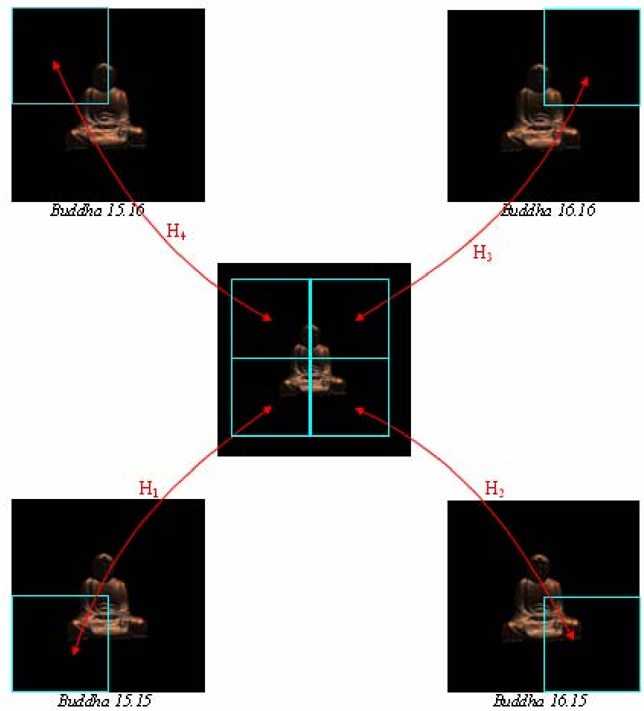

Fig. 7 The relation between the regions of the rendered image and the original light field images. H_1, H_2, H_3 and H_4 correspond to the homographies between the illustrated regions.

In the detection process, first of all, the rendered image should be partitioned into regions according the original light field images from which the pixels in each region are generated. This partition process will be explained in detail in the next subsection. After such a partition of the rendered image, the homographies between the regions of the rendered image and corresponding regions of the light field images are found by means of utilizing the scale invariant feature points [13]. Then, the estimated homography relations are applied to the corresponding regions in the original watermark, W, to generate the rendered watermark, W_{ren} . Finally, the normalized correlation between the rendered watermark and the rendered image that has been passed from high pass filtering is computed.

The normalized correlations for the original watermark and randomly generated 100 watermark pattern are shown in Fig. 8 for the rendered image shown in Fig. 7. The rendered image corresponds to approximately 70-80 % scaled version of the original light field image. The watermark is successfully detected. Another example is shown Fig. 9 and 10. In this case, a rotation is also included in the imagery camera orientation. The watermark is detected successfully again.

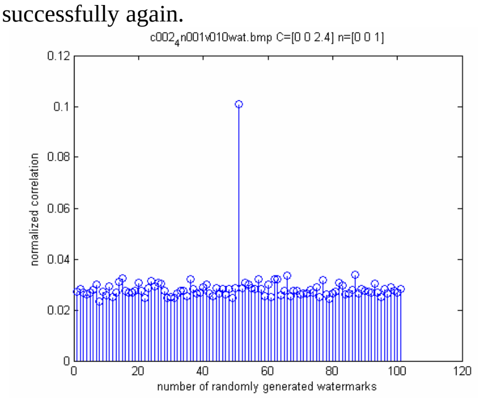

Fig. 8 Normalized Correlations results. Camera position = [0 0 2.4]. Image normal = [0 0 1]

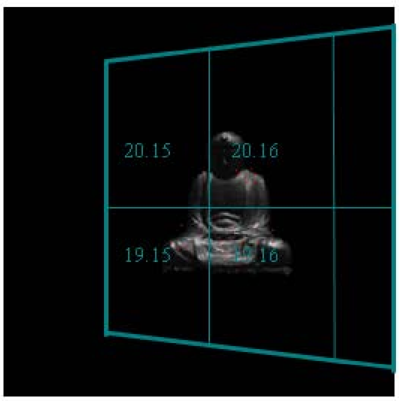

Fig. 9 The rendered image for the imagery camera at location [0.5 0 2.4] with orientation [0.3 0 1]. The numbers show the original image from which corresponding region is generated.

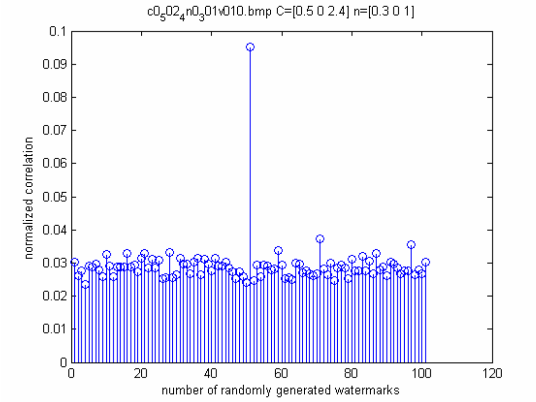

Fig. 10 Normalized Correlations results for the rendered image in Fig. 9

3.4 Decomposing the Rendered Image into Regions

In the third case given above, it is assumed that which region of the rendered image is generated from which original light field image is known (Fig. 7). However, in the actual case, the original light field images that are used in the generation of the rendered image should also be determined in a complete solution. For this purpose, an algorithm that exploits the invariant image metrics under projective planar transformation as in the second case is developed. The algorithm first finds the matched feature points between each original image and the rendered image. In Fig. 11, the matched points between the rendered image and the image buddha.15.15 are shown as an example. While the matched points in the square illustrated in Fig. 11.b originate from the image *buddha.15.15*, the other feature points belong to the other original light field images. Therefore, these points can be selected among other feature points by utilizing the projective invariance between the corresponding region in square and the image *buddha.15.15*. The algorithm to find the feature points coming from different light field images is given in [15].

The result of the algorithm for the rendered image generated for an imagery camera located at [0 0 2.4] with [0 0 1] orientation and the image buddha.15.15 is shown in Fig. 11.c. The selected feature points by the algorithm are mostly located on the region generated from the image buddha.15.15.

(a) *Rendered Image* **(b)** *Buddha.15.15*

(c) *found feature points*

Fig. 11 (a), (b) Matched feature points in the rendered image and the image *Buddha.15.15.* **(c)** The found matched points on the rendered image by the proposed algorithm.

After determination of the original light field image of each feature point, the homographies between the rendered image and light field images are estimated by using the matched points. In order to assign each pixel to a region, the found homographies are applied to the pixel coordinates and the difference between the intensity values of the pixel and its corresponding match is calculated. The pixel is assigned to the original light field image whose homography gives the minimum intensity difference. In order to increase the performance of the algorithm, the intensity difference is calculated within a 3x3 window for each pixel.

The results of the partition algorithm for the rendered image generated for an imagery camera located at [0 0 2.4] with [0 0 1] orientation are shown in Fig. 12. The regions of the % 95 of all pixels are correctly determined. Another example for the rendered image generated for an imagery camera located at [0.5 0 2.4] with [0.3 0 1] orientation is shown in Fig. 13. The regions of the % 90 of all pixels are correctly determined. The normalized correlation for the original watermark decreases by %5-10 due to imperfect region identification. However, this decrease is tolerable when the watermark detection performances illustrated in Fig. 8 and Fig. 10 are considered.

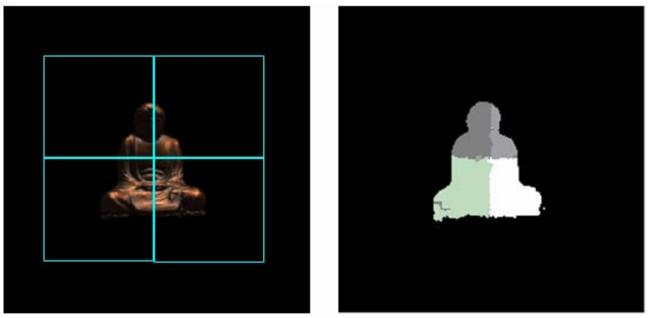

Fig. 12 The results of the partition algorithm.

4. CONCLUSIONS

A novel problem, called *free-view watermarking*, is introduced. The specific problems in imperceptibility and robustness requirements for *free-view watermarking* are presented. Assuming that the position and rotation for the imagery view is known, the proposed method extracts the watermark successfully from an arbitrarily rendered image. In order to extend the method for the case of an unknown imagery camera position and rotation, the variations on the watermark pattern during nearest neighborhood

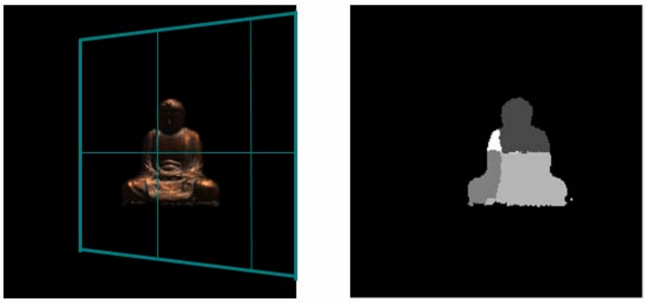

Fig. 13 The results of the partition algorithm.

interpolation are analyzed. Based on the analysis, the relation between the original watermark pattern and the watermark component on the rendered image is found as different forms of projective planar transformation. The solution of the problem is achieved by utilizing the found projective planar transformation. The embedded watermark is detected successfully for any unknown imagery camera position and orientation.

5. REFERENCES

[1] Alper Koz, Cevahir Çığla and A. Aydın Alatan, “Free view Watermarking for Free view Television”, IEEE ICIP, Atlanta, GA, USA, 8-11 October 2006.
[2] Cha Zhang, Tsuhan Chen, “A survey on image based rendering-representation, sampling and compression”, Signal Processing: Image Communications, 19 (2004) 1-28.
[3] A Survey report on state-of-the-art technologies for representing the 3D Content of a Scene to be utilized in 3D TC applications. European Union Project. Project Number: 511568, Project Acronym: 3DTV.
[4] C. Fehn, et al, “Interactive virtual view video for immersive TV applications”, Proceedings of IBC ’01, Amsterdam, Netherlands, September 2001, Vol. 2, pp. 53–62.
[5] Masayuki Tanimoto, “FTV (Free Viewpoint Television) Creating Ray-Based Image Engineering”, ICIP, Genova, September 11-14, 2005.
[6] Ingemar J. Cox et al, “Secure Spread Spectrum Watermarking for Multimedia”, IEEE Trans. on Image Processing,1673-1687, (1997).
[7] A. Koz and A. A. Alatan, “Oblivious Video Watermarking Using Temporal Sensitivity of HVS”, IEEE ICIP, Volume 1, 11-14 Sept. 2005, Page(s):961 - 964
[8] Gerhard C. et al, “Watermarking Digital Image and Video Data”, IEEE Signal Processing Magazine, September 2000.
[9] M. Levoy, P. Hanrahan, “Light field rendering”, (SIGGRAPH’96), New Orleans, LA, August 1996, pp. 31–42.
[10] http://graphics.stanford.edu/software/lightpack/lifs.html
[11] http://amp.ece.cmu.edu/projects/MobileCamArray/
[12] Richard Hartley and Andrew Zisserman, “Multiple View Geometry in Computer Vision”, Cambridge University Press.
[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," IJCV, 60, 2 (2004), pp. 91-110.
[14] Maes et al, “Digital Watermarking for DVD Video Copy Protection”, IEEE Signal Processing Magazine, September 2000.
[15] Alper Koz, Cevahir Cigla and A. Aydin Alatan, “Free View Watermarking”, submitted to IEEE Transactions on Forensics and Information Security.