Multi-view Object Representation with Modified 2-Layer IDP Decomposition

ROUMEN KOUNTCHEV
Radio-communications Dept.
Technical University of Sofia
Bul. Kl. Ohridsky, 8
Sofia 1000, BULGARIA
rkountch@tu-sofia.bg
http://www.tu-sofia.bg

VLADIMIR TODOROV
T&K Engineering Co.
Mladost 3, POB 12
Sofia 1712
BULGARIA
todorov_vl@yahoo.com

ROUMIANA KOUNTCHEVA
T&K Engineering Co.
Mladost 3, POB 12
Sofia 1712
BULGARIA
kountcheva_r@yahoo.com

Abstract: In the paper is presented one new method for multi-view object representation based on image decomposition with modified Inverse Difference Pyramid. The method offers new approach for efficient description of the multi-view images using one of them as a reference one. The decomposition has a relatively low computational complexity because it is based on orthogonal transforms (Walsh-Hadamard, DCT, etc.). The relations which exist between transform coefficients from the consecutive decomposition layers permit significant reduction of the coefficients needed for the high-quality object representation.

Key-Words: Multi-view images, multi-view representation, image pyramid decomposition

1 Introduction

Within the pattern-recognition and computer-vision the problem of defining representative multi-views for recognition and representation of 3D objects has recently received significant attention [1, 2, 3, 4].

Scientists and industry increasingly need multi-view representations of objects in the built environment and the demand for such kind of information is ever increasing. Some of the typical application areas are:

- 3D geographical information systems (GIS);
- hazardous and accident site survey;
- quality control for production lines;
- facility or construction management;
- object data mining, etc.

Two different types of image features can be extracted: those that are directly related to the 3D shape of the part of the object being viewed and features, that result from the 3D to 2D down projection (the second one can be ambiguous because part of the 3D shape information is lost during the projection). The essence of the recognition problem is to relate the structures found in the image with the underlying object models. Other important issues involved in structural recognition are:

- ➢ adequacy of the representation for the kind of objects encountered;
- ➢ selection and extraction of visual primitives;
- ➢ description of the spatial relations between primitives;
- ➢ matching image structures to models;
- ➢ inference of structural object descriptions from examples.

The selection of suitable shape primitives is of central importance. For efficient recognition, they should be *expressive* in the sense that a combination of only a few of them, or even a single one, can facilitate object identification. On the other hand, the available shape primitives should be *general* enough to model a large range of object categories. They should be detectable from images reliably in a bottom-up fashion and should be *non-incidental*, in the sense that they are unlikely to occur from random configurations in space [5, 6].

The pyramidal image representation is one of the frequently used techniques. The object reconstruction at a given pyramid level is based on the feature-based matching approach. The first step required at each level is the extraction of salient features (points and/or lines) together with their topological relations, which is a process controlled by a model of what is expected to be found in the images. Having detected features in two or more images, the correspondence problem has to be solved. The general approach [7, 8, 9] seeks correspondences in object space, because this approach is more flexible with regard to handling occlusions and surface discontinuities. The task-dependent local model of the object surface is then provided, and false correspondences are detected from bad fits to that model in object space [10, 11].

The Inverse Difference Pyramid (IDP) decomposition [12, 13] suits the peculiarities of this basic approach. The IDP-based object representation (and correspondingly - the salient features extraction) is performed in the spectrum domain. The creation of consecutive approximating images with increasing quality suits very well the widely used algorithms for image data mining [11]. Together with this the IDP decomposition offers specific advantages when the creation of 3D object model is concerned.

The paper is arranged as follows: in Section 2 are given the principles of the multi-view image representation with modified 2-layer Inverse Difference Pyramid (IDP) Decomposition; in Section 3 are presented some experimental results and Section 4 is the Conclusion.

2 Multi-view object representation based on 2-layer IDP

The IDP–based approach for multi-view representation is similar with the human way for object analysis and recognition [6] – starting with low similarity and continuing with better resemblance. By analogy, the object model creation starts with the coarse approximation, which corresponds to the lower pyramid layer, and continues with better approximation obtained in the next one(s). The object model is completed when the required quality for the best approximation is obtained.

The multi-view image representation of a still object is obtained by processing the object views taken from various positions. The selection of the view points positions and their number depends on the application. The general approach is the view points to be placed in a sphere, surrounding the object (i.e. the object is in the center of the sphere). For some applications, the view points could be arranged in a line, or in a circle. An example multi-view arrangement in a part of a sphere is shown in Fig. 1 and typical arrangement in a circle – in Fig. 2. The optimum number of view points (correspondingly – the angles between them) depends on the application as well. One of the views is always used as a reference one.

For example, if the needed multi-view should represent objects on a theatre scene, the view points should be placed in a plane and their number should correspond to the seats in the simulated hall. In this case, the view points should be placed in a relatively small sector of a sphere. Else, if for example, the application is to represent objects in the way they are seen by an insect, the view points number and positions should be quite larger.

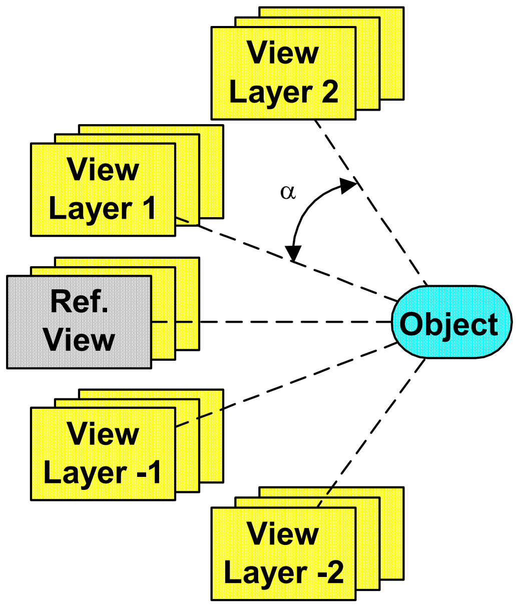

Fig. 1. Example layered multi-view arrangement in parallel circles, which build a part of a sphere around the object.

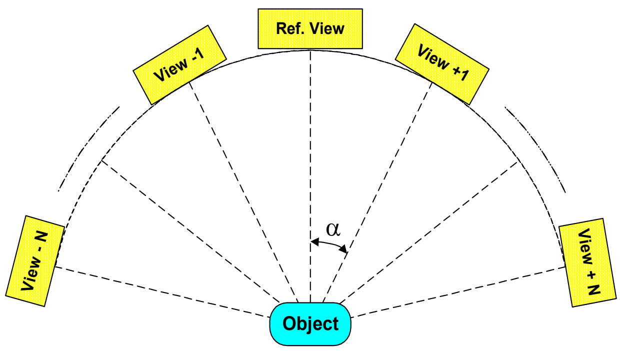

Fig. 2. View points arranged on an arc

The algorithm for the multi-view object representation based on 2-layer IDP is presented below.

For the creation of the multi-view representation of an object in a scene is necessary to prepare the initial (2N+1) multi-view images. For this, (for the case when the view points are arranged in a plane) are used (2N+1) cameras, placed at regular distances of same angle, α :

$$\alpha = \frac{\varphi}{2N+1}, \qquad (1)$$

on a part of a circle, where in the center of the circle is placed the object. The angle φ defines the width of the viewing angle, which (for practical applications) is usually chosen to be in the range of 20^0 and 30^0.

The object multi-views are processed individually. Each image is represented as a digital matrix, [B]. The IDP decomposition is “truncated”, i.e. the image is initially divided into sub-blocks (sub-images), which are after that processed sequentially, in accordance with the arrangement, shown in Fig. 3 below.

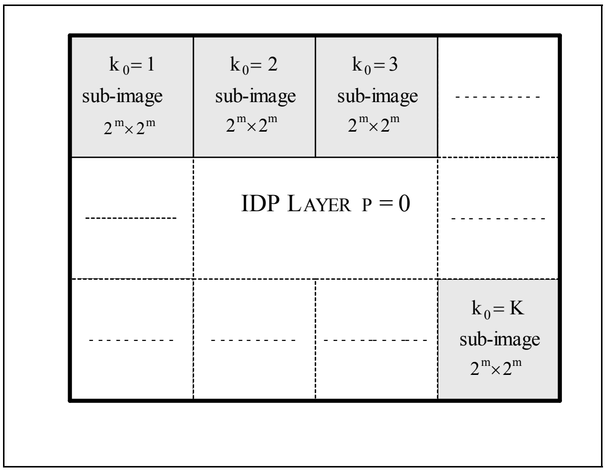

Fig. 3. Sub-blocks arrangement of the image matrix [B] in the decomposition layer p = 0

Each image sub-block which is a part of the n^{th} multi-view image of the object, is represented by the matrix $[B_n]$ for N = 0, ± 1, ±2, . . , ±N of size $2^m\times 2^m$. The matrix $[B_0]$ corresponds to the so-called “reference” image, placed at the middle of the sequence of consecutive views $[B_n]$ for n = 0.

In order to make the information redundancy in the sequence of matrices $[B_n]$ for n=0, ±1, ±2, .. , ±N smaller, in this work is offered to use the following modification of the image IDP decomposition of 2 layers:

1. For the IDP layer p = 0 is calculated the transform $[S_0^0]$ of the reference image $[B_0]$ by applying the direct orthogonal transform, as follows:

$$[S_0^0]=[T_0][B_0][T_0], \tag{2}$$

where $[T_0]$ is the matrix of the selected 2D direct orthogonal transform of size $2^m\times 2^m$.

2. The matrix of the approximated transform of the reference image is calculated:

$$[\hat{S}_0^0]=[m_0(u,v)s_0^0(u,v)], \tag{3}$$

where $m_0(u,v)$ is the element of the matrix-mask $[M_0]$ used to set the retained spectrum coefficients:

$$m_0(u,v)=\begin{cases}1, \text{ if } s_0^0(u,v) \text{ - retained coefficient,}\\ 0 \quad\quad - \quad\quad \text{ in all other cases,}\end{cases} \tag{4}$$

3. The first approximation $[\hat{B}_0]$ of the reference image is calculated performing the inverse orthogonal transform:

$$[\hat{B}_0]=[T_0]^t[\hat{S}_0^0][T_0]^t, \tag{5}$$

where $[T_0]^t = [T_0]^{-1}$ is the matrix of the inverse orthogonal transform of size $2^m\times 2^m$.

4. The difference matrix is calculated:

$$[E_0]=[B_0]-[\hat{B}_0]. \tag{6}$$

5. The so-obtained difference (error) matrix is divided into 4 sub-matrices:

$$[E_0]=\begin{bmatrix}[E_0^1] & [E_0^2]\\ [E_0^3] & [E_0^4]\end{bmatrix}, \tag{7}$$

where $[E_0^i]$ for i = 1, 2, 3, 4 are sub-matrices of size $2^{m-1}\times 2^{m-1}$.

6. For the IDP layer p = 1 is calculated the transform $[S_0^i]$ of the i^{th} sub-matrix of the difference $[E_0]$, using direct orthogonal transform:

$$[S_0^i]=[T_1][E_0^i][T_1] \text{ for } i=1,2,3,4, \tag{8}$$

where $[T_1]$ is the matrix of the direct orthogonal transform, of size $2^{m-1}\times 2^{m-1}$. The error matrix division into sub-matrices for the next decomposition layer is shown on Fig.4.

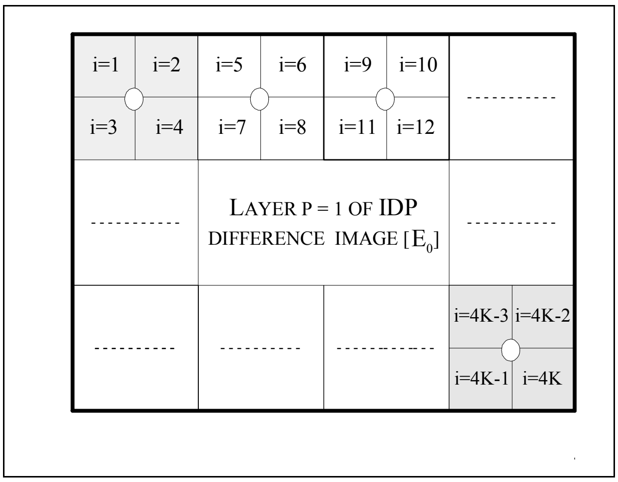

Fig.4. Sub-blocks arrangement in the difference image $[E_0]$ in the decomposition layer p = 1.

7. The approximated i^{th} transform is then calculated:

$$[\hat{S}_0^i] = [m_1(u,v)s_0^i(u,v)], \qquad (9)$$

where $m_1(u,v)$ is an element of the matrix-mask $[M_1]$ used to define the retained transform coefficients:

$$m_1(u,v)=\begin{cases} 1, \text{if } s_0^i(u,v) \text{ - retained coefficient,} \\ 0 \quad - \quad \text{in all other cases,} \end{cases} \qquad (10)$$

8. For the decomposition layer p=1 of the view $[B_n]$ is calculated the difference:

$$[E_n] = [B_n] - [\hat{B}_0] \text{ for } n=0,\pm1,\pm2,..,\pm N. \qquad (11)$$

9. The difference matrix is divided into 4 sub-matrices:

$$[E_n] = \begin{bmatrix} [E_n^1] & [E_n^2] \\ [E_n^3] & [E_n^4] \end{bmatrix}. \qquad (12)$$

where $[E_n^i]$ for i=1,2,3,4 are sub-matrices of size $2^{m-1}\times2^{m-1}$.

10. The i^{th} transform $[S_n^i]$ of the sub-matrix of the difference $[E_n^i]$ is obtained using direct orthogonal transform:

$$[S_n^i] = [T_1][E_n^i][T_1] \text{ for i=1,2,3,4.} \qquad (13)$$

11. The approximated i^{th} transform is calculated (in fact, this is the spectrum of the difference matrix $[E_n^i]$):

$$[\hat{S}_0^i] = [m_1(u,v)s_0^i(u,v)], \qquad (14)$$

where $m_1(u,v)$ is an element of the matrix-mask, used to define the retained transform coefficients:

$$m_1(u,v)=\begin{cases} 1, \text{if } s_0^i(u,v) \text{ - retained coefficient,} \\ 0 \quad - \quad \text{in all other cases,} \end{cases} \qquad (15)$$

12. The difference matrices of the approximated transforms are calculated:

$$[\Delta\hat{S}_n^i] = [\hat{S}_0^i] - [\hat{S}_n^i] \text{ for } n = \pm1, \pm2,.., \pm N. \qquad (16)$$

13. The values of the coefficients of the spectrum matrices $[\hat{S}_0^0]$ and $[\Delta\hat{S}_0^i]$ for i=1,2,3,4 and n = 0, ±1, ±2, ... , ±N in the decomposition layers p=0,1 are losslessly coded.

Eqs. (11) and (16) represent the basic difference between the basic IDP decomposition, and the modification, used for the multi-view processing. In the basic IDP decomposition each image has its own approximation for the consecutive decomposition layers. Besides, in the modified approach, presented here, all views use the same coarse approximation, and in the second approximation only, each view is processed individually. Another significant difference is that the basic IDP decomposition usually comprises 3 or 4 layers, starting with large image sub-blocks. The modification, used for the multi-view representation, is based on 2-layer decomposition, built for relatively small sub-blocks, usually of size 8 x 8 pixels for the lower layer and 4 x 4 pixels – for the higher one.

The decoding is performed in reverse order and is represented by the operations given below:

1. The coefficients of matrices $[\hat{S}_0^0]$ and $[\Delta\hat{S}_0^i]$ for i = 1, 2, 3, 4 and n = ±1, ±2, . . , ±N in the decomposition layers p = 0, 1 are decoded;

2. The approximated transforms in the decomposition layer p = 1 are restored:

$$[\hat{S}_n^i] = [\hat{S}_0^i] + [\Delta\hat{S}_n^i] \text{ for } n = \pm1,\pm2,..,\pm N. \qquad (17)$$

3. For the reference image (n = 0) is calculated each i^{th} approximated sub-matrix $[\hat{E}_0^i]$ of the difference matrix $[\hat{E}_0]$, applying inverse 2D orthogonal transform:

$$[\hat{E}_0^i] = [T_1]^t[\hat{S}_0^i][T_1]^t \text{ for } i = 1,2,3,4. \qquad (18)$$

4. For the decomposition layer p = 0, is calculated the matrix of the approximated reference image $[\hat{B}_0]$, applying the corresponding inverse orthogonal transform:

$$[\hat{B}_0] = [T_0]^t[\hat{S}_0^0][T_0]^t, \qquad (19)$$

5. The matrix $[\hat{B}]$ of the restored reference image is calculated (n = 0):

$$[\hat{B}] = [\hat{B}_0] + [\hat{E}_0]. \qquad (20)$$

6. The difference sub-matrices $[\hat{E}_n^i]$ for i=1,2,3,4 of the multi-view images in the decomposition layer p=1 are calculated, using the correspondent inverse orthogonal transform:

$$[\hat{E}_n^i] = [T_1]^t[\hat{S}_n^i][T_1]^t \text{ for } n = \pm1, \pm2,.., \pm N. \qquad (21)$$

7. The matrices $[\hat{B}_n]$ of the restored multi-views are calculated:

$$[\hat{B}_n] = [\hat{E}_0] + [\hat{B}_0] \text{ for } n = \pm1, \pm2,.., \pm N, \qquad (22)$$

The matrices $[\hat{B}_n]$ of all image blocks of size $2^m\times2^m$, which are involved in the multi-view images for $n = 0, \pm 1, \pm 2, \ldots, \pm N$ are decoded in accordance with the method, described above.

The difference between the basic IDP decomposition, and the modification, used for the multi-view processing in the decoding, is represented by Eqs. (17) and (22). Here, the restoration of the reference view image is performed in the way it is done in the basic IDP decomposition, i.e. the two approximations are used directly for the creation of the reference view image, and the remaining views in the same sequence are restored using the coarse approximation for the reference image and the fine approximation, belonging to the corresponding view.

The block diagram of the coder for multi-view object representation based on 2-layer IDP decomposition is shown in Fig. 5.a, and the block diagram of the decoder - on Fig. 5.b. The two block diagrams correspond to the methods for coding and decoding of grayscale multi-view images with modified IDP decomposition, described above. The two block diagrams represent the processing of one sub-block of the processed image. The coding of color multi-view images is performed in similar way, but it requires the color components to be processed individually. Depending on the color format (RGB, YUV, YC_RC_B, KLT, etc.), and the color sampling format (4:4:4, 4:2:0, 4:1:1, etc.) for each component is built individual pyramid. The approach based on the processing of the reference image and the remaining ones in the group, is retained.

The processing of multi-view images obtained from cameras, arranged in a part of a sphere, is performed in similar way.

3 Experimental results

For higher efficiency the approach presented here is based on the use of a fixed set of transform coefficients (these of lowest spatial frequency). For the experiments was used a 2-layer decomposition. In the low decomposition layers a set of 4 coefficients is usually enough. In the last (highest) layer is possible to use one coefficient only, which results in more efficient description efficiency. For the experiments was used IDP with the Walsh-Hadamard orthogonal transform (WHT). The views were obtained by moving the photo camera in a line, with an angle of 4^o between every two adjoining view positions. The total number of views in a line was 11. The reference image was chosen to be the one in the middle of the sequence. Two more view lines (11 views each) were arranged by moving the photo camera 4^0 up and down in correspondence to the first. The processed images were of size 864×576 pixels, 24 bpp each. The reference image from one of the test groups is shown in Fig. 6.

The same experiments were performed using DCT instead of the Walsh-Hadamard orthogonal transform. The results were similar: the restired images quality was a little higher (with about 0,2 dB) but the compression (i.e. the representation efficiency) was lower (with about 0,5). Taking into account the lower computational complexity of the WHT transform, in this paper are given the results obtained for the WHT transform.

Fig. 6. The reference view for Test sequence 1.

The example objects are convex and this permits relatively small number of views to be used for their representation. The experimental results for the first line of test images are given in Table 1 below.

Table 1. Results for the first line of consecutive views

View No.	CR Layer 2	L2 file size [B]	PSNR L2 [dB]	CR (group)
Ref.	181,45	6 171	36,83	69,16
1	89,16	12 560	35,55	87,44
2	99,29	11 277	36,25	98,60
3	110,25	10 157	36,54	107,44
4	118,00	9 490	36,58	114,72
5	133,89	8 363	36,53	121,98
6	129,53	8 645	36,51	127,33
7	117,51	9 529	36,45	130,38
8	107,43	10 423	36,24	131,69
9	100,53	11 138	36,37	131,93
10	92,47	12 110	35,81	131,09
Mean PSNR = 36,32 dB				

All experiments were performed transforming the original RGB images into YC_RC_B with sampling format 4:2:0.

In Fig. 7 are shown the first (a) and the last (b) image in one of the test sequences. The angle between the first and the last view is 20^0.

Fig. 7.a. The first image in the Test sequence 2.

Fig. 7.b. The last image in the Test sequence 2.

Despite the apparent similarity between the images, corresponding to the two views placed at the ends of the processed sequence of image views, the difference between them is large (Fig. 8).

Fig. 8. Difference between the first and the last image in Test sequence 2.

Similar example is given in Figs. 9 and 10, which represent view images and the difference between the first and the last one for Test sequence 3.

For the experiments, the basic sub-image in the low decomposition layer was 8 × 8 pixels and the number of the low-frequency transform coefficients was set to be 4 (the retained coefficients correspond to the low-frequency 2D Walsh-Hadamard functions). The size of the coarse approximation file (layer 1) for the reference view was 15418 B and the corresponding PSNR was 37,83 dB. The mean PSNR for the whole group of 11 views for 2-layer IDP was 36,32 dB. The compression ratio was calculated in accordance with Eq. 23:

$$CR = \frac{(2N+1)r^2 b_0}{(L_0 + 2NL_1)b_s}, \tag{23}$$

where b_0 и b_s represent the number of bits for one pixel and one transform coefficient correspondingly; L_0 and L_1 – the number of the retained coefficients for the IDP layers p = 0 and p = 1.

The so defined compression ratio does not represent the influence of the lossless coding of the coefficients' values performed for IDP layers p = 0 and p = 1.

In the column "L2 file size" is given the size of the corresponding approximations for the higher decomposition layer in Bytes. The compression ratio (CR) was calculated for the whole group of images, i.e. the total data needed for the representation of all 11 views was compared with the uncompressed data for the same images. In the column named "CR Layer 2" is given the compression ratio obtained for the corresponding representations of the decomposition layer 2 only.

Similar investigation was performed for another 11 views of the same objects, placed in a line positioned at 4^0 higher than the first one. The angles between adjacent views were 4^0. In this case the reference view was chosen to be at the end of the sequence (next to View No. 10). The results are given in Table 2.

Table 2. Results obtained for the second line of consecutive views (4^0 up) for Test sequence 1.

View No.	CR Layer 2	L2 file size [B]	PSNR L2 [dB]	CR (group)
Ref.	90,62	16 746	35,65	129,98
1	70,80	21 088	34,47	126,32
2	75,00	19 906	35,56	123,99
3	80,75	18 490	34,53	122,74
4	82,23	18 157	34,52	121,83
5	86,17	17 326	34,53	121,40
6	90,26	16 541	34,61	121,34
7	89,36	16 708	34,70	121,23
8	88,04	16 959	34,81	121,03
9	86,99	17 162	35,11	120,78
10	85,16	17 532	35,26	120,43
Mean PSNR = 34,89 dB				

Fig. 9.a. The first image in the Test sequence 3

Fig. 9.b. The last image in the Test sequence 3.

Fig. 10. Difference between the first and the last image in Test sequence 3.

The results obtained are close to those given in Table 1, but the CR and the PSNR are a little lower, because the reference view for the second line was set to be this at the end of the sequence and as a result, the correlation between the consecutive views is lower.

Additional test was performed for a line of consecutive views positioned at 4^0 down in respect of the first one. The global results are as follows: the PSNR for the whole group (3 lines of views) is 34,8dB and the compression ratio is CR = 120,1. This means that for the group of 33 color images (one reference image and 32 views arranged in 3 adjoining lines) each of size 864 × 576 pixels, was achieved a compression ratio CR > 120. The quality of the views was visually lossless, because the errors in images which have a PSNR higher than 32 dB are imperceptible (Fig. 11).

The tests performed simulated a matrix of 33 views arranged in a rectangle of size 11 × 3. Best results were obtained for the case, when the reference view was placed in the center of the viewing matrix.

The main advantage of the new approach is that it ensures high compression and very good quality of the restored visual information. In spite of the global approach when multi-view data storage is concerned, each view could be restored individually and used. Compared with the famous JPEG standard for still image compression, the method offers much higher quality of the restored images. For example, the mean PSNR of an image after compression 100:1 is 24,6 dB, but the visual quality of the restored image is very bad (Fig. 12).

Fig. 11. Restored reference image after IDP compression 100:1.

Fig. 11. Restored image after JPEG compression 100:1.

The image from Fig. 7a, compressed with JPEG2000-based software gave for same compression a result image with PSNR = 34,4 dB (a little lower than that, obtained for the reference image with the new method), but the computational complexity of JPEG 2000 is much higher and the background of the image was visually woollier.

For a group of images, comprising all multi-views (the test sequences, used for the investigation), comparison was not done, because JPEG2000 does not offer similar option and the results should be just a sum from all views, i.e. there is no cumulative effect.

Additional disadvantage is that JPEG 2000 does not offer the ability for retained coefficients reduction, which is possible when the IDP decomposition is used, because of the specific relations between the coefficients' values in neighboring decomposition layers.

4 Conclusion

In the paper is presented one new method for multi-view object representation based on modified 2-layer IDP image decomposition. The method ensures very efficient description of the multi-view images by using one of them as a reference one. The decomposition has a relatively low computational complexity because it is based on orthogonal transforms (DCT, Walsh-Hadamard, etc.). For example, the computational complexity of decompositions, based on wavelet transforms is much higher. The comparison of the computational complexity of the IDP decomposition and the wavelets-based transforms is given in earlier publications of the authors [12]. In the examples was used the WH transform, but DCT or some other transforms are suitable as well. The relations existing between transform coefficients from the consecutive decomposition layers permit significant reduction of the coefficients needed for the high-quality object representation [14]. The number of the necessary views depends on the application. For example, the view area could be restricted to some angle or scale, etc.

The experimental results proved the ability to create efficient multi-view object representation based on the IDP decomposition. The task is easier when the image of a single object has to be represented. In the examples, presented here, two convex objects were represented and they should be searched together. The significant compression of the data representing the multiple views ensures efficient data storage and together with this - fast access and search in large image databases.

The IDP representation is suitable for tasks requiring the analysis of complicated scenes (several objects searched together or context-based search). This is possible, because the lowest layer of the pyramidal decomposition consist of sub-images, processed individually. In result, more than one object (described individually) could be searched together.

Additional advantage is the similarity of the transform coefficients from any two adjacent decomposition layers, which is a basis for the creation of flexible algorithms for the transformation of the already created object representation into higher or lower scale without using additional views.

The future development of the method will be aimed at its application for video sequences analysis and moving objects representation [3, 15].

Acknowledgement
This work was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science, Contract VU-I 305.

References:

[1] T. Denton, M. Demirci, J. Abrahamson, A. Shokoufandeh, and S. Dickinson. Selecting canonical views for view-based 3-D object recognition. *In 17th International Conf. on Pattern Recognition* (ICPR'04), volume 2, August 2004, pp. 273-276.

[2] P. M. Hall and M. J. Owen. Simple canonical views. *In The British Machine Vision Conf.* (BMVC'05), volume 1, 2005, pp. 7-16.

[3] S. Kim, S. Lee and Y. Ho. Three-dimensional natural video system based on layered representation of depth maps. *IEEE Trans. On Consumer Electronics*, Vol. 52, No. 3, Aug. 2006, pp. 1035-1042

[4] F. Mokhtarian and S. Abbasi. Automatic selection of optimal views in multi-view object recognition. In *The British Machine Vision Conf.* (BMVC'00), 2000, pp 272-281.

[5] P. Schyns and A. Oliva. From blobs to boundary edges: evidence for time and spatial scale dependent scene recognition, *Psychol. Sci.* No 5, 1994, pp.195-200.

[6] D. Hubel. Eye, Brain and Vision, *Scientific American Library*, No 22, W. Freeman, NY, USA, 1989.

[7] S. Takahashi, I. Fujishiro, Y. Takeshima, and T. Nishita. A feature-driven approach to locating optimal viewpoints for volume

visualization. In *IEEE Visualization* 2005, pp. 495-502.
[8] J. Todd. The visual perception of 3D shape. *TRENDS in Cognitive Science*, 8(3): 2004, pp. 115-121.
[9] P. Vazquez, M. Feixas, M. Sbert, and W. Heidrich. Automatic view selection using viewpoint entropy and its applications to image-based modelling. *Computer Graphics Forum,* 22(4), 2003, pp. 689-700.
[10] A. Oliva and A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, *Int. J. Computer Vision* Vol. 42, 2001, pp. 145-175.
[11] W. Kropatsch, H. Bischof (Eds.), *Digital image analysis: selected techniques and applications*. Springer - Verlag, NY, Berlin, Heidelberg, 2001.
[12] R. Kountchev, M, Milanova, C. Ford, R. Kountcheva. Multi-layer Image Transmission with Inverse Pyramidal Decomposition. In: *Computational Intelligence for Modeling and Predictions*, S. Halgamuge, L. Wang (Eds.), Vol. 2, Chapter 13, Springer-Verlag Berlin, Heidelberg, 2005, pp. 179–196.

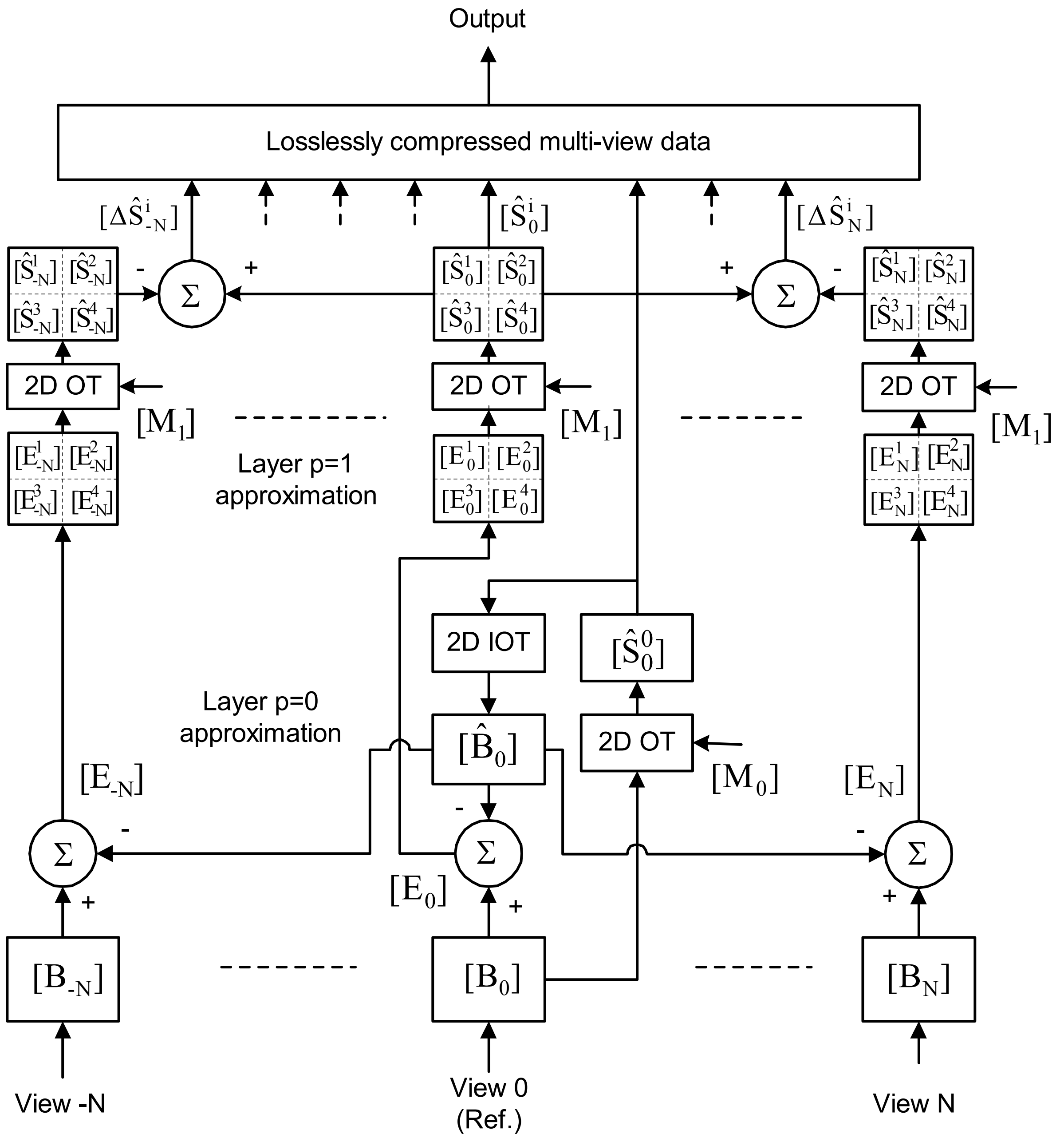

Fig. 5.a. Block diagram of the coder for multi-view object representation based on Modified 2-layer IDP decomposition

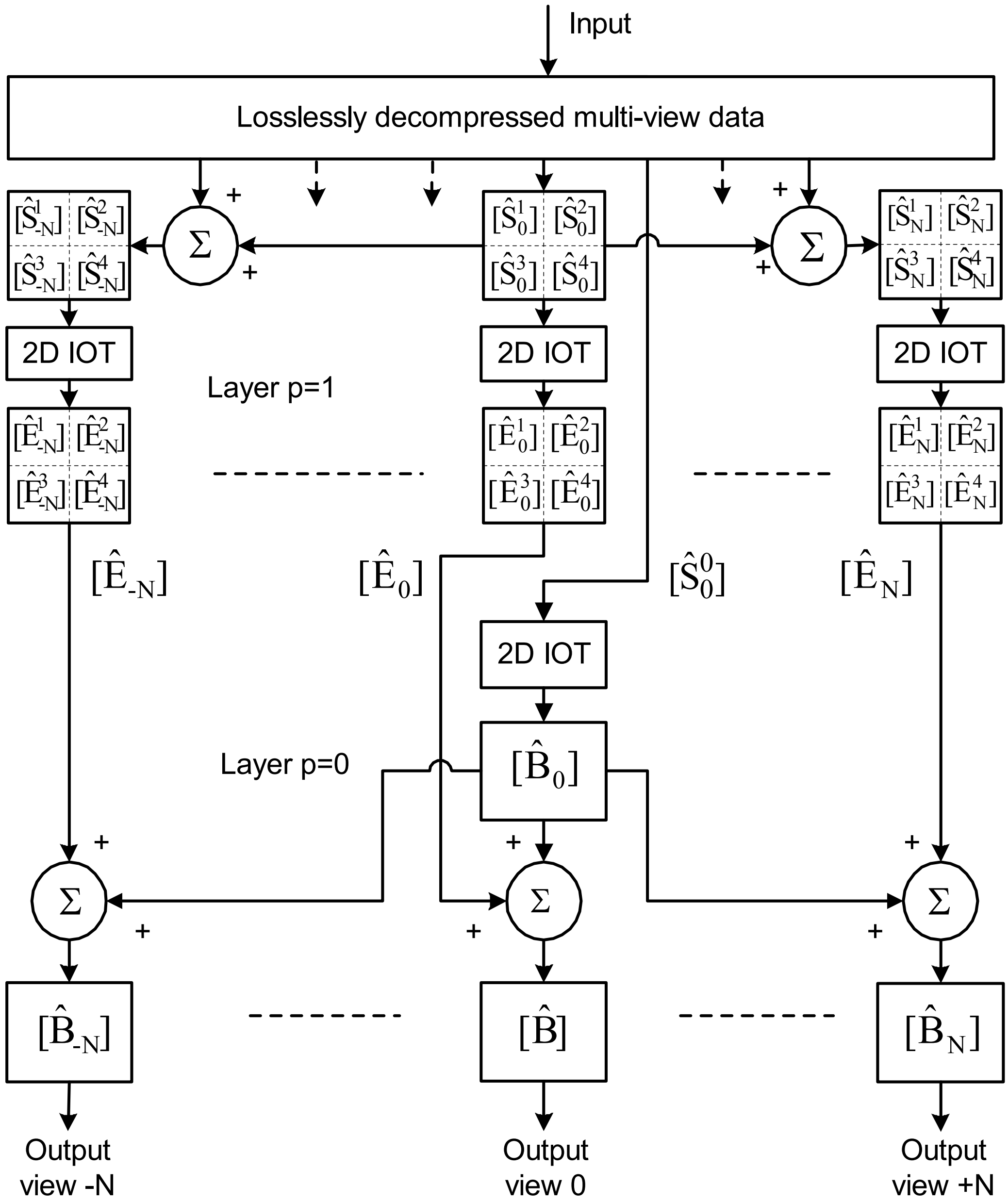

Fig. 5.b. Block diagram of the decoder for multi-view object representation based on Modified 2-layer IDP decomposition

Abbreviations: 2D OT – two-dimensional orthogonal transform; 2D IOT - two-dimensional inverse orthogonal transform.

[13] R. Kountchev, S. Rubin, M. Milanova, Vl. Todorov, R. Kountcheva. Cognitive Image Representation Based on Spectrum Pyramid Decomposition. *Proc. of the WSEAS Intern. Conf. on Mathematical Methods and Computational Techniques in Electrical Engineering* (MMACTEE'08), Sofia, Bulgaria, May 2-4, 2008, pp. 230-235.

[14] R. Kountchev, R. Kountcheva. Image Representation with Reduced Spectrum Pyramid. Book chapter in: *New Directions in Intelligent Interactive Multimedia*, Eds.: G. Tsihrintzis, M. Virvou, R. Howlett, L. Jain. Springer-Verlag, Berlin, Heidelberg, 2008, pp. 275-284.

[15] ISO/IEC JTC1/SC29/Wg11 m12542: Multi-view video coding based on lattice-like pyramid GOP structure, 2005.