

**Military Technical College
Kobry El-Kobbah,
Cairo, Egypt**



**7th International Conference
on Electrical Engineering
ICEENG 2010**

Evaluation Of Local Disparity Estimation For View Interpolation

By

Ahmed Saeed* Hussein A.Aly* Alaa Abdel Reheem** Fawzy Ibraheem

Abstract:

In this paper, we discuss the disparity estimation problem which is the most important challenge to generate an intermediate image for view synthesis. The main problem in estimating the disparity is to find the correspondence among the input images. The quality of the synthesized novel view is dependent on the accuracy of disparity estimation. Once we find the exact disparities among the input images, a simple low complexity interpolation is carried out to generate the new view. We offer a view of a comparative study between different state-of-the-art disparity estimation algorithms showing the advantages, disadvantages, and the peak signal to noise ratio (PSNR) of the reconstructed view using each one.

* Egyptian Armed Forces

** Modern Technical Institute, Cairo, Egypt

*** College of Engineering, Cairo University, Cairo, Egypt

1. Introduction:

Continuous generation of a view of a scene based on the user position and orientation is gaining research attention recently. It is the idea behind the new multiview 3D displays or free-view video technology [1]. This continuous generation of novel views can be accomplished by recording the light rays at every space location towards every possible direction over any range of wavelengths and at any time which is defined by the plenoptic function [2] that may require a huge storage which is impossible. Alternatively acquiring a fewer set of scene images from different views, the required virtual view can be interpolated for other viewing angles or positions [3], this is called the *view interpolation*. The key of success of the view interpolation is the accuracy of disparity estimation (how to relate different views by horizontal and vertical displacement).

The quality of the interpolated view mainly depends on how accurately the computed disparities. In [4], the authors divide the existing techniques for disparity estimation into four categories :

- Feature based techniques: in which a correspondence between features of the input images is established. The advantage of this technique is the reduction in time complexity, as the correspondence is done based on small amount of data not the whole image pixels. The drawback of this technique is that its poor accuracy as some regions of the input images are not included in the matching process.
- Area based techniques: in which the input images are divided into regions (patches), and then the correspondence between them is established as in [5]. The drawback of this technique is that it is very sensitive to the illumination changes and additive noise and the assumption that all points in a patch should have the same disparity does not always hold.
- Phase based techniques: in which the input images are matched based on the local phase information. They are more robust to different texture levels and less sensitive to brightness variations as in [6] and [7]. The computed disparity map suffers abrupt changes in flat areas, and distorted edges as the computed disparities are not smooth.
- Energy based techniques: in which the disparity estimation problem is done based on minimization of an objective function in a regularization framework. The problem is solved by discretizing the associated Euler-Lagrange equations for the functional minimization as in [8], [9], and [10]. The main drawback of this technique is that it is very sensitive to the selection of the regularization parameters as discussed later in

section 2.2.

In this paper, we will compare among the commonly used methods: area based, phase based, and energy based techniques in details as a local disparity estimation techniques. The global disparity estimation techniques such as [11] is not included in this paper as it is not suitable for view interpolation problem since the input images have multi-depth and local motions. Also we compare the reconstructed images using these different approaches both visually and quantitatively using PSNR, which serves as a guide to which technique can be used in solving the view interpolation problem for different situations.

The rest of this paper is organized as follows. Section 2 describes the algorithms used in our comparative study. In section 3, we present the experimental results. Finally, we give concluding remarks in section 4.

2. Algorithms under comparison:

In this paper, we use all discussed algorithms to reconstruct a new view between two input (left and right) images. We denote the left image as I^l and the right image as I^r and the reconstructed view as I^{rec} , which all are defined on a sampling grid \mathcal{G} .

$$\hat{\mathbf{d}}_k = \arg \min_{\mathbf{d}_k} U(I^l(k), I^r(k + \mathbf{d}_k)) \quad (1)$$

where $\hat{\mathbf{d}}_k$ is the estimated disparity for the token k , $k \in \{1, 2, 3, \dots, n\}$, n is the number of image tokens, and $U(\cdot)$ is some cost function. The token can be the whole image thus the disparity becomes only one vector for all the pixels in the input image, but this is not suitable for view interpolation case because there are many local motions in the scene.

Thus the token can either be :

- Block (block-based disparity estimation) which is one of the areas based techniques as discussed in section 2.1.
- Pixel (pixel-based disparity estimation) in which we select two energy minimization (isotropic and anisotropic regularization) and one phase based technique (Gabor transform) as discussed in section 2.2.

2.1 Block-Based Approach:

The token in equation (1) here is a block with 8×8 or 16×16 pixels (commonly used in MPEG standards), which divides I^f into macro-blocks. These macro-blocks are compared with a corresponding block and its adjacent neighbors within a window of p pixels on all four sides of the corresponding macro-block in I^l as shown in figure (1) (a). For large motions, we select p large enough to locate the block in the search window. A very important issue to note is that the block-based approach does not get the true motion for every frame; instead it locates the best match macro-block i.e. the macro-block that minimizes some cost function such as the mean absolute difference

$$(MAD) = \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} |I^l(m, n) - I^r(m, n)|$$

or mean squared error

$$(MSE) = \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} (I^l(m, n) - I^r(m, n))^2$$

where N is the macro-block

side length (8 or 16). The search algorithm affects the speed and accuracy of the block matching algorithm. For instance, if we search all macro-blocks within the search window (exhaustive search), we get the best accuracy but the highest computational burden. On the other hand, inaccurate fast search methods such as three-step-search used in MPEG encoding have a poor accuracy to be used for view interpolation. In general, the computation can be speeded up by implementing the block matching in a multi-resolution framework [13] by estimating the disparity field in the coarse resolution and then refining it in the next higher resolution after scaling it by appropriate scaling factor as shown in figure (1) (b).

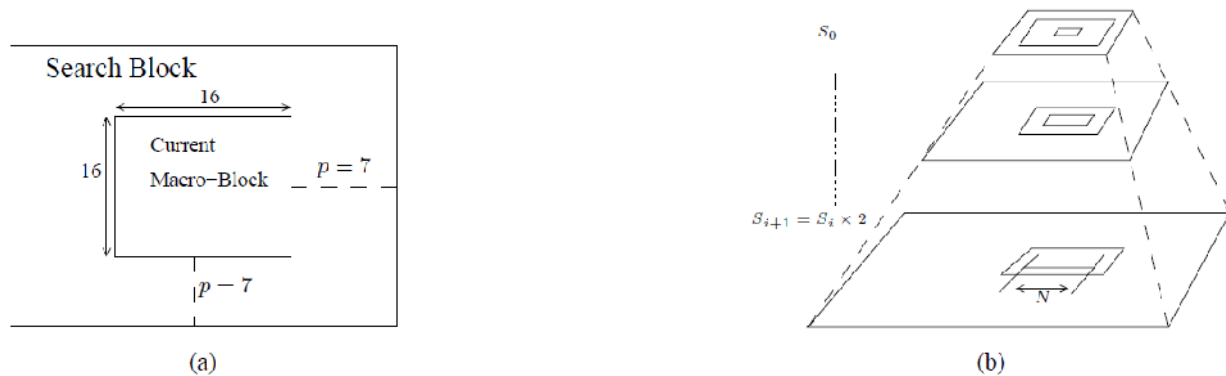


Figure (1): (a) A typical example of macro-block size and p parameter, (b) Illustration of multi-resolution block matching algorithm

2.2 Pixel-Based Approach

The token here that relate I^l to I^r is the pixel. Let $d = [u, v]^T$ be the disparity vector that originates at spatial location (x, y) at I^r and pointing to I^l as shown in figure (2). To get d the following assumption must hold:

$$I^r(x, y) \approx I^l(x + u, y + v) \quad (2)$$

Figure (2): disparity estimation using pixel-based approach. $I^l(x, y)$ is the left image and $I^r(x, y)$ is the right image.

This paper For each pixel we want to determine both values of u and v that minimize the difference between both sides of equation (2), since we have only one equation, this problem is called ill-posed problem. Mathematically, it needs additional constraints to be solved properly. We select two energy minimization based techniques which are the isotropic and anisotropic regularization for our discussion. In addition to the Gabor transform as one phase-based technique.

2.2.1 Isotropic regularization

The unconstrained problem described in (2) is solved using additional constraint which is the similarity of the disparity vectors of the neighboring pixels as in [8]. So it takes the following form

$$d(x) = \min_{\hat{d}(x)} \int \int_{x \in \Gamma} \left(I^l(x + \hat{d}(x)) - I^r(x) \right)^2 + \lambda \left[\|\nabla_x u\|^2 + \|\nabla_x v\|^2 \right] dx \quad (3)$$

where ∇_x is the spatial gradient operator, $(I^l(x + \hat{d}(x)) - I^r(x))^2$ is called the data term, and $(\|\nabla_x u\|^2 + \|\nabla_x v\|^2)$ is the regularization term and λ is a parameter that control the degree of smoothness. From the calculus of variation, the solution of the minimization problem in (3) can be done by solving its associated Euler Lagrange equations as follows:

$$\frac{\partial U}{\partial u} = \left(I^1(\mathbf{x} + \hat{\mathbf{d}}(\mathbf{x})) - I^r(\mathbf{x}) \right) I_x^1(\mathbf{x} + \hat{\mathbf{d}}(\mathbf{x})) - \lambda [u_{xx} + u_{yy}] = 0, \quad (4)$$

and

$$\frac{\partial U}{\partial v} = \left(I^1(\mathbf{x} + \hat{\mathbf{d}}(\mathbf{x})) - I^r(\mathbf{x}) \right) I_y^1(\mathbf{x} + \hat{\mathbf{d}}(\mathbf{x})) - \lambda [v_{xx} + v_{yy}] = 0, \quad (5)$$

where I_x^1 and I_y^1 are the spatial derivative of I in both x and y directions respectively, and $u_{xx}, u_{yy}, v_{xx}, v_{yy}$ are the second derivative of u and v with respect to x and y respectively. Equations (4) and (5) can be discretized using gradient descent method as follows:

$$\frac{\partial U}{\partial u} = \frac{u^{n+1} - u^n}{\Delta t}, \quad (6)$$

$$\frac{\partial U}{\partial v} = \frac{v^{n+1} - v^n}{\Delta t},$$

where u^n and v^n are the computed u and v at iteration n respectively and Δt is an artificial time-step parameter to ensure the stability of the solution. Then we solve iteratively

$$u^{n+1} = u^n - \Delta t \left(\frac{\partial U}{\partial u} \right), \quad (7)$$

$$v^{n+1} = v^n - \Delta t \left(\frac{\partial U}{\partial v} \right),$$

The disadvantage of the selection of such regularization term is the smoothness of the disparity vectors done regardless to the object boundaries which result in some distortion at the object boundaries.

2.2.2 Anisotropic regularization

The idea introduced by Porona and Malik [14] can be used to solve the smoothness problem of the disparity vectors across the object boundaries resulting from the isotropic regularization. This is achieved by introducing the anisotropic diffusion which diffuses the intensities of the image pixels inside the objects, and stops this smoothness across the object boundaries. Many authors [9] used this idea and applied it to the estimation of disparity vectors instead of image intensities as shown in the following equation

$$\mathbf{d}(\mathbf{x}) = \min_{\hat{\mathbf{d}}(\mathbf{x})} \int \int_{\mathbf{x} \in \Gamma} \left(I^l(\mathbf{x} + \hat{\mathbf{d}}(\mathbf{x})) - I^r(\mathbf{x}) \right)^2 + \lambda \zeta(I^l(\mathbf{x}), u, v) d\mathbf{x}, \quad (8)$$

where

$$\zeta(I(\mathbf{x}), u, v) = \lambda \left[\frac{1}{(1 + I_x^2(\mathbf{x}))^2} (u_x(\mathbf{x})^2 + v_x(\mathbf{x})^2) + \frac{1}{(1 + I_y^2(\mathbf{x}))^2} (u_y(\mathbf{x})^2 + v_y(\mathbf{x})^2) \right]$$

and u_x, u_y, v_x, v_y are the derivatives of u and v with respect to x and y respectively. Both terms $\frac{1}{(1+I_x^2(\mathbf{x}))^2}$ and $\frac{1}{(1+I_y^2(\mathbf{x}))^2}$ are used to guide the smoothness process thus the disparity will be smoothed in the direction of the lowest image gradient and preserved in the direction of the highest image gradient. The solution of the minimization problem in (8) can be done by solving its associated Euler-Lagrange equations as follows:

$$\frac{\partial U}{\partial u} = \left(I^l(\mathbf{x} + \hat{\mathbf{d}}(\mathbf{x})) - I^r(\mathbf{x}) \right) I_x^l(\mathbf{x} + \hat{\mathbf{d}}(\mathbf{x})) - \lambda \left\{ \frac{\partial}{\partial x} \left[\frac{u_x(\mathbf{x})}{(1 + (I_x^l)^2(\mathbf{x}))^2} \right] + \frac{\partial}{\partial y} \left[\frac{u_y(\mathbf{x})}{(1 + (I_y^l)^2(\mathbf{x}))^2} \right] \right\} = 0, \quad (9)$$

$$\frac{\partial U}{\partial v} = \left(I^l(\mathbf{x} + \hat{\mathbf{d}}(\mathbf{x})) - I^r(\mathbf{x}) \right) I_y^l(\mathbf{x} + \hat{\mathbf{d}}(\mathbf{x})) - \lambda \left\{ \frac{\partial}{\partial x} \left[\frac{v_x(\mathbf{x})}{(1 + (I_x^l)^2(\mathbf{x}))^2} \right] + \frac{\partial}{\partial y} \left[\frac{v_y(\mathbf{x})}{(1 + (I_y^l)^2(\mathbf{x}))^2} \right] \right\} = 0, \quad (10)$$

Equations (9) and (10) can be solved by the gradient descent method as in (7). Since the gradient of underlying image is the guide to the diffusion process, the anisotropic diffusion fails in some cases, when there is a weak gradient between the object and the background. This represents the main weakness of the anisotropic diffusion (edge preserving regularization). The main drawback of all regularization techniques is that the reconstructed image is sensitive to the regularization parameter which need to be chosen according to the contents of I^r and I^l .

2.2.3 Gabor-transform

The Gabor transform is implemented using a set of modulated limited-band Gaussian filters at different tuning frequencies and orientations determined by the horizontal and vertical spatial frequencies ω_x, ω_y respectively. Its 2D unit impulse is given by

$$g(x, y) = \frac{1}{2\sigma_x\sigma_y} e^{-\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)/2} \times e^{[2\pi j(\omega_x x + \omega_y y)]} \quad (11)$$

where σ_x and σ_y are the standard deviations of the prototype Gaussian filter in both x and y directions respectively. We choose σ_x and σ_y such that the modulating central frequencies are $\{1/16, 1/8, 1/4\}$ and at four orientations $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ as in [15], obtaining the 12 filter kernels applied to I^r and I^l . Let G_j^i be the filtered I^i image by the j^{th} Gabor filter, $1 \leq j \leq 12$. The disparity field d is estimated by comparing every pixel at $x \in [x_1, x_2]$ in $G_j^l(x)$ to its correspondent and all the pixels in a window in $G_j^r(x)$ by summing the absolute values of their differences as in [15]. We assume that the horizontal disparity component $u \in [x_1, x_2]$ and the vertical disparity component $v \in [y_1, y_2]$, and the disparity estimated by :

$$\mathbf{d}(\mathbf{x}) = \arg \min_{\hat{\mathbf{d}}(\mathbf{x})} \sum_j \left| G_j^r(\mathbf{x} + \hat{\mathbf{d}}(\mathbf{x})) - G_j^l(\mathbf{x}) \right|^2 \quad (12)$$

The selection of values of x_1, x_2, y_1 and y_2 is dependent on the largest horizontal and vertical displacements between the left and right images.

3. Experimental results

We apply the discussed four techniques to some natural sequences to test the effectiveness of each one. In all experiments, we interpolate the new image at distance $\alpha = 0.5$ using the estimated disparity field after scaling it by α . Since we have the complete sequence, we measure the PSNR between the ground truth images with the interpolated one. In figures (3) and (4) (a) is I^l , (b) is I^r , (c) is the ground truth image, and (d) is I^{rec} using 8×8 block size. The PSNR of the reconstructed image is 22.93db and 24.84db for figures (3) and (4) respectively. This is due to the fact that the images in figure (3) have much more complex objects than the other figure. From these figures, we find that the block based approach suffer from the following two problems :

1. The blockiness artifacts (distortion of the reconstructed image at object boundaries) This can be reduced by reducing the size of the block size but to a limit that the matching is done properly, or first calculates the disparity values for larger blocks and then reduces the block size at possible boundary locations as

proposed in [5].

2. We notice that some parts are visible in the left image and disappeared in the right image, and vice versa, these parts are called occluded and newly exposed area respectively. The matching process does not know what to do in these regions but it gets the block with the smallest cost in the search window. We surround some parts of the reconstructed images by circles. These parts are reconstructed poorly because of the occluded and newly exposed areas in I^l and I^r which make the PSNR of I^{rec} falls down.

Figure (3): Sample disparity estimation for the natural Flower-Garden sequence that have local motions using block based method with block size 8×8 . (a) is I^l , (b) is I^r , (c) is the ground truth image, and (d) is I^{rec} using 8×8 block size. The PSNR of the reconstructed image is 22.93db

Another example for pixel-based disparity estimation is shown in figure (5), in which (a) is I^l , (b) is I^r , (c) is the ground truth image, (d) is I^{rec} using isotropic regularization, (e) is I^{rec} using anisotropic regularization, (f) is the computed disparity map using isotropic regularization, and (g) is the computed disparity map using anisotropic regularization. The computed disparity map using isotropic regularization suffers from the problem of over smoothness of disparity vectors which is greatly reduced using the anisotropic regularization that shows sharper objects boundaries thus obtaining more

accurate result. We employ a 4-level hierarchical implementation in order to avoid local minima, and bicubic interpolation to estimate sub pixel intensities. We set $t = 1.5 \times 10^{-5}$ and $\lambda = 2000$ in all our experimental results as in [9]. The PSNR between the reconstructed and ground truth image is 25.93db and 26.4db using isotropic and anisotropic regularization respectively, i.e. the anisotropic regularization gets 0.5db gain over the isotropic regularization. In the same figure (5) (h) is I^{rec} by the disparity estimation using Gabor transform as in (12), and figure (5) (i) is the computed disparity map. The PSNR between the reconstructed and

Figure (3): Sample disparity estimation for the natural Calendar-Train sequence that have local motions using block based method with block size 8×8 . (a) is I^l , (b) is I^r , (c) is the ground truth image, and (d) is I^{rec} using 8×8 block size. The PSNR of the reconstructed image is 24.84db.

Figure (5): Pixel-based disparity estimation methods : (a) is I^l , (b) is I^r , (c) is the ground truth image , (d) is I^{rec} using isotropic regularization, (e) is I^{rec} using anisotropic regularization, (f) is the computed disparity map using isotropic regularization, (g) is the computed disparity map using anisotropic regularization. (h)

is I^{rec} by the disparity estimation using Gabor transform, and (i) is the compute disparity map using Gabor transform. The PSNR between the reconstructed and ground truth image is 25.93db using isotropic regularization and 26.4db using anisotropic regularization and 23.5db using Gabor transform.

Figure (6): Plot the PSNR between the reconstructed and the ground truth images against the frame number of the Flower-Garden sequence. Disparity between every two successive frame with difference two (I_t, I_{t+2}) of the sequence is computed. We take I_t as the I^l and I_{t+2} as I^r and the ground truth image as I_{t+1}

ground truth image is 23.5db using Gabor transform. From a coarse point of view, the disparity estimated by the Gabor transform method is good. However, there are still some obvious errors in the disparity map, like some noisy values or abrupt changes in flat areas, or some distorted edges, as can be seen from figure (5)(i). To alleviate such

errors we need to get the disparity values for continuous surfaces changing smoothly, while maintaining the disparity discontinuities at the object boundaries using the anisotropic diffusion discussed previously. So, the disparity map computed by the Gabor transform method serves as initial estimate and must be refined by a one of a variational regularization method as in [15].

Finally, we present a summary of measured PSNR between the ground truth image and the I^{rec} using the discussed four disparity estimation techniques. We use 30 frames from the Flower-Garden sequence to get such unbiased results. We compute the disparities between every two successive frame with difference two (I_t, I_{t+2}). We take I_t as the I^l and I_{t+2} as I^r and the ground truth image as I_{t+1} . In figure (6), we plot the PSNR between the reconstructed and the ground truth images against the frame number of the sequence for all methods.

4. Conclusion

In this paper, we present different disparity estimation techniques comparing their reconstructed images visually and quantitatively. We show that the block-based technique has the lowest quality in terms of the reconstructed image since it suffers from the blockiness artifacts and the wrongly handling occlusion areas as shown in figure (3) and (4). The pixel based methods shows superior performance than the block based methods as shown in figure (6). Also the reconstructed image by the regularization methods (isotropic, and anisotropic) has a better PSNR than those obtained by Gabor transform method since the disparity vectors obtained by Gabor transform method is not smooth (not regularized). Regularization make the disparity estimation problem more constrained and enforces the disparity vectors to be smooth which handles occlusion areas much better, which offers a better solution than those that do not have any regularization. The advantage of estimating disparity using Gabor transform is that it can handle large disparities [15] which are not possible for energy-based techniques.

References:

- [1] Neil A. Dodgson, "Autostereoscopic 3D Displays," *Computer*, vol. 38, pp. 31–36, 2005.
- [2] M. Landy and J. Anthony Movshon, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, E.H. Adelson and J.R. Bergen, Eds., chapter 1, pp. 3–21. The MIT Press, Cambridge, MA, 1991.
- [3] Takeo Kanade, Peter Rander, and P.J. Narayanan, "Virtualized Reality: Constructing Virtual Worlds from Real Scenes," *IEEE MultiMedia*, vol. 4, pp. 34–47, 1997.
- [4] Javier Sanchez, Rachid Deriche, Luis Alvarez, and Joachim Weickert, "Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scale-Space Based Approach," Tech. Rep., INRIA Sophia Antipolis, January 2000.
- [5] Mancini, and J. Konrad, "Robust quadtree-based disparity estimation for the reconstruction of intermediate stereoscopic images," *Proceedings SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 3295, pp. 53–64, 1998.
- [6] D.J. Fleet, "Disparity from local weighted phase correlation," *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, vol. 1, pp. 48–56, 1994.
- [7] X. Huang and E. Dubois, "Disparity estimation for the intermediate view interpolation of stereoscopic images," *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing (ICASSP 05)*, vol. 2, pp. 881–884, Mar. 2005.
- [8] J. Konrad, "Enhancement of viewer comfort in stereoscopic viewing: parallax adjustment," *Proceedings SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 3639, pp. 179–190, 1999.
- [9] S. Ince and J. Konrad, "Occlusion-Aware View Interpolation," *EURASIP Journal on Image and Video Processing*, vol. 2008, Oct. 2008, Article ID 803231, 15 pages.
- [10] S.B. Kang, R. Szeliski, and J. Chai, "Handling occlusion in dense multi-view stereo," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, vol. 1, pp. 103–110, Dec 2001.
- [11] Siavash Zokai and George Wolberg, "Image Registration Using Log-Polar Mappings for Recovery of Large-Scale Similarity and Projective Transformations," *IEEE transaction on image processing*, October 2005.
- [12] Anthoni Mancini, "Disparity estimation and intermediate view reconstruction for novel applications in stereoscopic video," M.S. thesis, Department of Electrical Engineering, McGill University, 1998.

- [13] B. Liu and A. Zaccarin, “*New fast algorithms for the estimation of block motion vectors*,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 3, pp. 148–157, April 1993.
- [14] P. Perona and J. Malik, “*Scale-space and edge detection using anisotropic diffusion*,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, pp. 629–639, Apr. 1990.
- [15] X. Huang and E. Dubois, “*Two-dimensional regularized disparity estimation based on the Gabor transform*,” Visual Communications and Image Processing (San Jose, CA), SPIE, vol. 6077, pp. 342–351, Jan. 2006.