IEIE Transactions on Smart Processing and Computing, vol. 7, no. 1, February 2018
https://doi.org/10.5573/IEIESPC.2018.7.1.049

Depth Map Estimation based on Maximum a Posteriori Probability

Olgierd Stankiewicz* and Marek Domański

Chair of Multimedia Telecommunications and Microelectronics, Poznań University of Technology / Poznań, Poland

***** Corresponding Author: Olgierd Stankiewicz, ostank@multimedia.edu.pl

Received November 16, 2017; Accepted December 11, 2017; Published February 28, 2018

** Regular Paper*

***Abstract*:** In this paper, we consider depth map estimation expressed as an optimization problem. We focus on the fitness function, for which we present a theoretical derivation based on a maximum a posteriori probability (MAP) rule. This is then used to show relations of interest with commonly used similarity metrics: sum of absolute differences (SAD) and sum of squared differences (SSD). The original derivations are also used to propose a depth estimation method. The experimental results are obtained with an implementation made on the basis of Moving Picture Expert Group (MPEG) Depth Estimation Reference Software (DERS). We show that with the proposed approach, it is possible to estimate a depth that allows higher quality of view synthesis (up to 2.8 dB of PSNR – Peak Signal-to-Noise Ratio) versus the original unsupervised DERS, when sub-optimal control parameters are used. If the DERS control parameters are optimized manually, the attained gain is smaller (up to 0.08 dB PSNR) but still does not need manual selection of control parameters.

***Keywords*:** Depth map estimation, Global optimization, Maximum a Posteriori probability, Graph cuts, Belief propagation, SAD, SSD

1. Introduction

Stereoscopic depth is used in fields like computer vision/graphics and three-dimensional (3D) video. In the latter, depth in the form of depth maps is used, along with multiview video, in order to create a representation of a 3D scene. Such a representation, called multiview video plus depth (MVD), has recently gained interest among researchers working on delivery formats for autostereoscopic displays, free viewpoint television (FTV), and 3D video compression. Therefore high-quality depth maps are needed for production of the content, including experimental material.

There are many ways to obtain depth maps, but all suffer from some weaknesses. For example, in natural scenes, depth maps can be acquired with the use of special depth-sensing cameras. Unfortunately, their practical usability is mostly restricted to indoor scenes due to limited measurement ranges, interference between cameras, environmental restrictions, etc.

A more general solution consists of the algorithmic estimation of depth maps from multiple views, e.g. from stereoscopic pairs. Although many solutions are known, algorithmic estimation of depth is still a demanding task with respect to both the quality of the estimated depth and the computational complexity of the algorithms. Moreover, for practical reasons, depth estimation is expected to be automatic, without the need for human intervention. Unfortunately, a variety of state-of-the-art depth estimation algorithms are controlled with parameters that have to be selected manually, because there are no models allowing automatic selection. Moreover, there is often a lack of theoretical views on depth estimation, which disallows development of such models. Therefore, the idea in this paper is to provide a theoretical approach to depth estimation based on maximum a posteriori probability (MAP) to cope with the mentioned problems.

2. State of the Art

Depth map estimation is most often performed by a dense search for correspondence between multiple views (at least two). The correspondence found for each pixel, expressed as the disparity between views, can easily be used to calculate depth if the baseline of the camera system

is known [1].

Disparity/depth estimation can be expressed as an optimization problem, which can then be solved with the use of generic methods. The goal is to find the global optimum (not a local optimum), and therefore, the energy function *Fitness* over a depth map is defined as

$$Fitness = \sum_{P} FitCost_{P} \tag{1}$$

where $FitCost_p$ depicts a sub-component of the *Fitness* function for particular pixel p in the considered disparity map. Such a function is often referred to as *energy*, *goal function*, or *performance index* in other fields.

Because such a *Fitness* function is formulated on a per-pixel basis, it can be used in a variety of generic optimization algorithms. Among the many algorithms known (such as genetic optimization), only a few of them have found applications in the field of depth map estimation due to the fact that the number of considered disparity values is relatively large (i.e. in the hundreds). The most commonly used optimization algorithms are graph cuts (GC) and belief propagation (BP) [2]. However, the descriptions of those algorithms are outside the scope of this paper, as those are used solely as tools for optimization of depth maps with regard to the *FitCost* function.

The term $FitCost_P$ is typically modeled as a sum of two sub-functions: *DataCost* and *TransitionCost*, for each pixel:

$$FitCost_P = DataCost_p\left(d_q\right) + \sum_{q \in N(p)} TransitionCost_{p \to q}\left(d_p, d_q\right), \tag{2}$$

where

- p – the pixel (point) for which *FitCost* is evaluated
- d_p – the assumed disparity of pixel p
- q – some pixel (point) in neighborhood $N(p)$ of pixel p
- d_q – the assumed disparity of pixel q

$DataCost_p\left(d_p\right)$ models the direct correspondence between pixels, and expresses how given pixel p is similar to those pointed to by disparity d_p in other images.

$TransitionCost_{p,q}\left(d_p, d_q\right)$ penalizes disparity maps that are not smooth. If given pixel p has a vastly different disparity, d_p, than its neighbors (pixels depicted by q), it gets a high *TransitionCost* penalty.

Of course, more advanced approaches than the one presented in (2) are known [2-7], where a higher-order *FitCost* function is defined, but their application is not very common [10, 11]. For example, an algorithm implemented in the Depth Estimation Reference Software (DERS) [24] developed within the ISO/IEC Moving Pictures Expert Group (MPEG), which is widely used in the literature, is a reference for comparison of different depth estimation algorithms, and employs formulation (2).

The usage of *DataCost* and *TransitionCost* is a common idea in all global optimization methods like belief propagation or graph cuts. Depending on the approach, those are defined as probabilities [12-15] or in terms of energy [2, 6, 16]. Li [17] used the mathematic concept of the partition function, related to Boltzmann probability distribution, in order to change an energy formulation into a probability, and vice versa. Unfortunately, there is a lack of empirical verification as to whether such operations are justified. This lack is one of the motivations for this paper. We have introduced the overall idea of this paper in [18] but here we present it in more details along with enhanced results.

2.1 *DataCost* Function

The DataCost function models the direct correspondence between pixels, and expresses how given pixel p is similar to those pointed to by its disparity, d_p, in other images. The higher the difference between those pixels, the higher the value of $DataCost_p\left(d_p\right)$.

The most commonly used *DataCost* is defined in terms of energy related to similarity metrics between fragments of images, calculated in pixels or blocks. Typically, the sum of absolute differences (SAD) [19] or the sum of squared differences (SSD) [14, 20] metrics are used. Some of the state-of-the-art work that relates to the *DataCost* function proposes the usage of *rank* or *census* [21] for calculation of better similarity metrics. Work [22] proposes a more advanced approach, where a mixture of various similarity metrics is used in order to obtain better depth estimation.

Sun et al. [14] provided a similar derivation of the *FitCost* function based on MAP assumptions. Unfortunately, their work focused mainly on a Gaussian model (corresponding to sum of squared differences energy formulation). Unfortunately, verification of whether such assumptions are correct was not provided.

Similarly, Cheng and Caelli [15] employed a posteriori probability for modeling of the *FitCost* function. A more advanced model for *DataCost* was considered, which incorporates a generalized Gaussian model with an arbitrary power exponent. Therefore, for an exponent value of 2, a Gaussian model was considered, and for a value of 1, a Laplace model was considered.

Zhang and Seitz [23] proposed usage of a truncated-linear *DataCost* function that actually responds to an absolute difference similarity metric but is limited so that it does not exceed a given maximal level.

Nevertheless, the above-mentioned works mainly do not deal with verification of the assumptions and they do not provide empirical data or theoretical analysis of the models.

Work [24] thoughtfully analyzed a probabilistic model of correspondence in 3D space. Instead of a MAP rule, a different approach to evaluating entropy and mutual information, called EMMA (EMpirical entropy

Manipulation and Analysis) was proposed. It is claimed that one of advantages of EMMA is that it does not require a prior model for the functional form of distribution of the data, and the entropy can be efficiently maximized (or minimized) using stochastic approximation. Nevertheless, the method was presented in the context of 3D modeling and not depth map estimation itself, which disallows comparison with the other state-of-the-art methods in the field.

2.2 Transition Cost Function

TransitionCost is a term of the *FitCost* function, which penalizes disparity maps that are not smooth. Its role is regularization of the resultant depth/disparity map. The higher the differences between disparity d_p of pixel p and disparity values d_q of all neighboring pixels d_q, the higher the value of $TransitionCost_{p\to q}\left(d_p, d_q\right)$.

Typically, $TransitionCost_{p\to q}\left(d_p, d_q\right)$ is defined independently from pixel positions p and q, and thus, it can be simplified to $TransitionCost\left(d_p, d_q\right)$. Also, very often, *TransitionCost* is not defined as a function of d_p and d_q independently, but as a function of $\left|d_p - d_q\right|$ only: $TransitionCost\left(\left|d_p - d_q\right|\right)$.

Among the most commonly known are three models for the *TransitionCost* function—the Potts model, the linear model, and the truncated-linear model.

a) Potts model [16]

$$TransitionCost\left(\left|d_p - d_q\right|\right) = \begin{cases} 0 & if \left|d_p - d_q\right| = 0 \\ \alpha & otherwise \end{cases} \quad (3)$$

b) Linear model [3, 25]

$$TransitionCost\left(\left|d_p - d_q\right|\right) = \gamma \cdot \left|d_p - d_q\right| \quad (4)$$

c) Truncated-linear model [23]

$$TransitionCost\left(\left|d_p - d_q\right|\right) = \min\left(\gamma \cdot \left|d_p - d_q\right| \;\; \alpha\right) \quad (5)$$

Used notation:

p – pixel for which *FitCost* function is evaluated

d_p – assumed disparity of pixel p

q – some pixel in the neighborhood of pixel p

d_q – assumed disparity of pixel q

α, γ – constant parameters

In general, *TransitionCost* functions incorporate some sort of constant parameter, like γ or α coefficients. The main purpose of such constant parameters is to provide weighting to the relation with the *DataCost* function, to which it is added to formulate *FitCost* function (1). The most commonly, parameter γ of the linear and truncated-linear models is called the *smoothing coefficient* (e.g. in MPEG DERS [25]), because its value sets how much any depth maps that are not smooth are penalized by the *FitCost* function. Usage of small values of the smoothing coefficient results in sharp depth maps, which are similar to those attained with local depth estimation methods. Usage of large values of the smoothing coefficient results in generation of very smooth, even blurred, depth maps. The selection of the smoothing coefficient is typically done manually (the depth estimation is thus supervised), which is an important problem in practical usage of depth estimation methods based on belief propagation or graph cuts in applications, where an unsupervised operation is expected.

All of the mentioned models (Potts, linear, and truncated-linear) are typically used because they are simple and provide some additional advantage in the case of a belief-propagation algorithm, because they allow reduction of computational complexity in the execution of particular steps, from an $O\left(D^2\right)$ polynomial to $O\left(D\right)$ linear time, where D is the number of disparity-considered values. As D typically ranges from 40 to 100, this provides a vast reduction in real computational complexity.

Zhang and Seitz [23] proposed the usage of a truncated-linear–shaped *TransitionCost* function for depth estimation and compared this against other state-of-the-art techniques. Although the results are promising, the foundations of the proposal were not given.

Papers [14] and [15] considered a derivation of the *TransitionCost* function based on a maximum a posteriori rule, similar to the approach in this paper. Based on this, a Markov random field model for stereoscopic depth estimation was formulated by means of a belief-propagation algorithm. Unfortunately, the work proposed only an approximation of the *TransitionCost* function.

The lack of research that provides theoretical analysis of the application of a maximum a posteriori probability optimization rule for the formulation of *DataCost* and *TransitionCost* for depth estimation, along with empirical experimentation that would support formulation of such theoretical models, is one of the motivations of this paper.

3. *DataCost* Derivation based on MAP

As mentioned in the introduction, one of the most crucial aspects in depth estimation is usage of pixel correspondence in the views. Based on similarity metrics between pixels, the best matching pixel pairs are chosen and used to derive disparity/depth.

In most of the work related to block matching (and depth estimation, in particular) no theoretical foundation is provided for the problem of optimal selection of the best match [14, 19, 20, 22, 25, 26]. Surprisingly, simple sum of absolute or squared differences similarity metrics (SAD or SSD in blocks) are often considered [14, 19, 20] without in-depth studies or consideration. Such an empirical approach, without theoretical formulation, is easy, but has disadvantages, as follows.

- It does not provide a scientific foundation for the

considerations.

- As there is no mathematical model, it is unknown if the obtained solution is the optimal one.
- Thus, it is difficult to incorporate empirical proposals as part of a broader framework, like optimization algorithms, where apart from the pixel similarity metric (referred to as *DataCost*), other terms are also used (e.g. *TransitionCost*).

Therefore, in this paper, a theoretical formulation based on MAP is derived.

Let us consider disparity estimation in the case of two cameras that are perfectly horizontally aligned with parallel optical axes. The views are rectified [27, 28], and the lens distortions [27, 28] are compensated. Therefore, epipolar lines are aligned with horizontal rows in the images.

Images from left view $L_{x,y}$ and right view $R_{x,y}$ have the same width, W, and the same height, H.

For a given row of pixels with coordinate y in both views, observed are pixel luminance values in the left view and the right view:

$L_{1,y}, L_{2,y} \ \ldots \ L_{W,y}$ – luminance values in the left view

$R_{1,y}, R_{2,y}, \ldots, R_{W,y}$ – luminance values in the right view (both indexed from 1 to W)

All of those are random variables, which are considered to have been already observed. Thus, these variables constitute our a posteriori observation set.

In depth estimation, for each pixel at coordinates x, y (in the right view), we search for the disparity value $d_{x,y}^{\ *}$ that would maximize the probability of $p\left(d_{x,y}\right)$ under the condition of a posteriori observation of luminance values in both views. This probability will be demarked $p_{x,y,d}$:

$$p_{x,y,d} \equiv p\left(d_{x,y} | \left(L_{1,y}, L_{2,y}, \ldots, L_{W,y}, R_{1,y}, R_{2,y}, \ldots, R_{W,y}\right)\right) \tag{6}$$

where $\left(L_{1,y}, L_{2,y} \ \ldots \ L_{W,y} \ R_{1,y} \ R_{2,y} \ \ldots \ R_{W,y}\right)$ is the overall conditional expression for observations of luminance values.

Therefore, a MAP rule for the search for optimal disparity value $d_{x,y}^{\ *}$ can be formulated as follows:

$$d_{x,y}^{\ *} = \underset{arg\ d}{\max}\left(p_{x,y,d}\right) \tag{7}$$

In order to allow the depth estimation algorithm to use MAP rule (7), the term $p_{x,y,d}$ has to be modeled based solely on values that are known after the observation (a posteriori), e.g. luminance values in left view $L_{1,y}, L_{2,y} \ \ldots \ L_{W,y}$ and in right view $R_{1,y}, R_{2,y}, \ldots, R_{W,y}$.

We will transform Eq. (6) using the Bayes rule:

$$p\left(A,B\right) = p\left(A\right) \cdot p\left(B|A\right) = p\left(B\right) \cdot p\left(A|B\right) \tag{8}$$

expressed in the form

$$p\left(B|A\right) = \frac{p\left(A|B\right) \cdot p\left(B\right)}{p\left(A\right)} \tag{9}$$

Thus, we get

$$p_{x,y,d} = \frac{p\left(\left(L_{1,y}, L_{2,y} \ \ldots \ L_{W,y} \ R_{1,y} \ R_{2,y} \ \ldots \ R_{W,y}\right) \ d_{x,y}\right) \cdot p\left(d_{x,y}\right)}{p\left(L_{1,y}, L_{2,y} \ \ldots \ L_{W,y} \ R_{1,y} \ R_{2,y} \ \ldots \ R_{W,y}\right)} \tag{10}$$

The expression for probability in the numerator of Eq. (10), $\left(\ldots\right) | \, d_{x,y}$, can be rearranged for each luminance separately and written as

$$\mathrm{p}\left(L_{1,y} \left| d_{x,y}, L_{2,y} \right| d_{x,y} \ \ldots \ L_{W,y}, \left| d_{x,y} \ R_{1,y} \right| d_{x,y} \ R_{2,y} \left| d_{x,y} \ldots \ R_{W,y} \right| d_{x,y}\right) \tag{11}$$

Assumed is the presence of noise that has independent realizations in each view. Therefore, each of the pixel luminance values in left view $L_{l,y}$ (at coordinates l, y) is independent from each of the pixel luminance values in right view $R_{r,y}$ (at coordinates r, y).

Moreover, when considering the denominator of Eq. (10), it can be assumed that pixel luminance values in left view $L_{1,y}, L_{2,y} \ \ldots \ L_{W,y}$ are also independent from each other, as are pixel luminance values in right view $R_{1,y}, R_{2,y}, \ldots, R_{W,y}$. Specifically, this also holds true for the sought pair of pixels matched by disparity $d_{x,y}$, because the denominator in Eq. (10) does not consider any specific matching or correspondence of pixels, as those probabilities are not conditional with respect to $d_{x,y}$. Therefore, we can simplify the denominator on the right side of Eq. (10) as follows:

$$\prod_{l=1..W} p\left(L_{l,y}\right) \cdot \prod_{r=1..W} p\left(R_{r,y}\right) \tag{12}$$

Similar simplification could also be done in the numerator of Eq. (10), simplified to Eq. (11), but here, on the contrary, probabilities of $L_{l,y} \, | \, d_{x,y}$ and $R_{r,y} \, | \, d_{x,y}$ are conditional, because they are considered under condition of the occurrence of $d_{x,y}$. Such a condition for $d_{x,y}$ means that in the given pixel with coordinates x, y, for which we calculate $p_{x,y,d}$, disparity value $d_{x,y}$ is assumed, so that two pixels in the left and right views correspond to each other. Such a pair of pixels is not independent, and therefore, probabilities of their luminance values, $p\left(L_{l,y}\right)$ and $p\left(R_{r,y}\right)$, cannot be simplified as in Expression (12). Such an exception occurs when coordinate l in the left view corresponds to the same pixel in the right view with coordinate r, which is true when l and r are linked by

disparity $d_{x,y}$:

$$r = x \quad ; \quad l = x + d_{x,y} \tag{13}$$

where (x expresses the coordinate in the right view for which $d_{x,y}$ is considered). For other pairs of pixels (not corresponding to each other), random variables describing their luminance values are independent, as in Expression (12). Therefore, we can express $p_{x,y,d}$ as

$$p_{x,y,d} = p\left(\left(L_{x+d_{x,y},y}, R_{x,y}\right) | d_{x,y}\right) \cdot p\left(d_{x,y}\right) \cdot \frac{\prod_{l=1..W, l \neq x+d_{x,y}} p(L_{l,y} | d_{x,y}) \cdot \prod_{r=1..W, r \neq x} p(R_{r,y} | d_{x,y})}{\prod_{l=1..W} p\left(L_{l,y}\right) \cdot \prod_{r=1..W} p\left(R_{r,y}\right)} \tag{14}$$

Also, with the exception for the mentioned case, Eq. (13), the probability distributions related to $p(L_{l,y} | d_{x,y})$ and $p(R_{r,y} | d_{x,y})$ are independent from $d_{x,y}$ (because those random variables represent pixels that are not connected by disparity $d_{x,y}$); thus, terms in the numerator of Eq. (14) can be simplified to:

$$\prod_{l=1..W, l \neq x+d_{x,y}} p\left(L_{l,y}\right) \cdot \prod_{r=1..W, r \neq x} p\left(R_{r,y}\right) \tag{15}$$

Now, we can see that all $\prod(\ldots)$ terms in the numerator of Eq. (15) can be simplified with $\prod(\ldots)$ terms in the denominator of Eq. (14). This applies to all l and r, except for Eq. (13). Thus, we can express $p_{x,y,d}$ as

$$p_{x,y,d} = \frac{1}{p\left(L_{x+d_{x,y},y}\right) \cdot p\left(R_{x,y}\right)} \cdot p\left(\left(L_{x+d_{x,y},y}, R_{x,y}\right) | d_{x,y}\right) \cdot p\left(d_{x,y}\right) \tag{16}$$

It can further be seen that the term $p\left(L_{x+d_{x,y},y}\right)$ is the probability distribution of luminance values in the left view, which is independent from corresponding disparity value $d_{x,y}$ and, therefore, can be expressed as $p\left(L_{x,y}\right)$. We finally get

$$p_{x,y,d} = \frac{1}{p\left(L_{x,y}\right) \cdot p\left(R_{x,y}\right)} \cdot p\left(\left(L_{x+d_{x,y},y}, R_{x,y}\right) | d_{x,y}\right) \cdot p\left(d_{x,y}\right) \tag{17}$$

The derivation of Formula (17) is one of the key achievements shown in this paper. It describes probability $p_{x,y,d}$ where the given pixel with coordinates x, y has disparity $d_{x,y}$ under the condition of the a posteriori observations of luminance values in both views.

Therefore, selection of $d_{x,y}$ that maximizes $p_{x,y,d}$ fulfills the MAP rule for Eq. (7). Later in the paper, it will be used in order to propose a novel depth estimation method.

3.1 Relations to SSD and SAD Similarity Metrics

In the meantime, we will show how Eq. (16) can be simplified in order to obtain classical squared differences (and thus, sum of squared differences for blocks—SSD) and absolute differences (and thus, the sum of absolute differences for blocks—SAD), pixel similarity metrics that are commonly used in depth estimation algorithms. The presented simplification is interesting, as it shows the set of conditions (resulting from assumptions) which, if met in a practical case, indicate that usage of SAD or SSD is optimal from a maximum a posteriori probability optimization point of view. Therefore, it will show in what cases usage of SAD or SSD is optimal. Note, though, that the presented reasoning does not limit the application of SAD or SSD pixel similarity metrics to the presented cases only.

Eq. (17) expresses probability $p_{x,y,d}$ that a given pixel with coordinates x, y has disparity $d_{x,y}$ based on the MAP rule. Terms $p\left(L_{x,y}\right)$ and $p\left(R_{x,y}\right)$ are probability distributions of luminance values in the left and right views, respectively. They can simply be measured as histograms of the left and right views. The interpretation of these terms is that correspondence between pixels with luminance values that occur more often is more probable. The mentioned terms are omitted by state-of-the-art pixel similarity metrics proposals. This corresponds to a situation where histograms of the compared images are flat.

Similarly, $p\left(d_{x,y}\right)$, the probability distribution of disparity values $d_{x,y}$, can be estimated as a histogram. It can be imagined that this brings some quality to the distinction between depth planes (e.g. foreground vs. background). This was also omitted from the state-of-the-art pixel similarity metrics proposals that correspond to situations where all disparities are equally probable.

The term $p\left(\left(L_{x+d_{x,y},y}, R_{x,y}\right) | d_{x,y}\right)$ is the probability that luminance value $L_{x+d_{x,y},y}$ of a pixel in the left view and luminance value $R_{x,y}$ of a pixel in the right view will occur, on the condition that those pixels correspond to each other, and the occurred disparity is i.

Again, according to Bayes rule in form (9), the term $p\left(L_{x+d_{x,y},y}, R_{x,y} | d_{x,y}\right)$ can be expressed alternatively as either

$$p\left(L_{x+d_{x,y},y}, R_{x,y} | d_{x,y}\right) = p\left(L_{x+d_{x,y},y}\right) \cdot p\left(R_{x,y} | L_{x+d_{x,y},y}, d_{x,y}\right) \tag{18}$$

or as

$$p\left(L_{x+d_{x,y},y}, R_{x,y} | d_{x,y}\right) = p\left(R_{x,y}\right) \cdot p\left(L_{x+d_{x,y},y} | R_{x,y}, d_{x,y}\right) \tag{19}$$

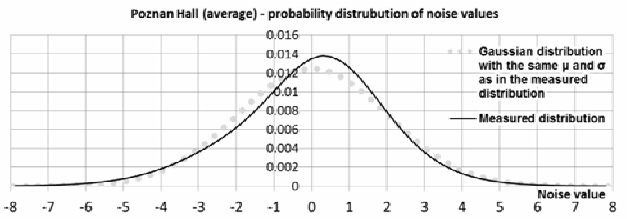


Fig. 1. Probability distribution of noise values in the Poznan Hall sequence (averaged over all views) measured in [29].

Those forms are equivalent, and lead to a similar formulation, so this work will only focus on the latter, Eq. (19). Term $p\left(R_{x,y}\right)$ simplifies the term in the denominator of Eq. (17):

$$p_{x,y,d} = \frac{p\left(d_{x,y}\right)}{p\left(L_{x,y}\right)} \cdot p\left(L_{x+d_{x,y},y} | R_{x,y}, d_{x,y}\right) \qquad (20)$$

In order to understand interpretation of the usage of the SAD or SSD similarity metric as a model for $p\left(L_{x+d_{x,y},y} | R_{x,y}, d_{x,y}\right)$, we have to make the following assumptions.

- The presence of additive noise, the same in both of the views (in particular, with equal standard deviation σ); below, we will consider two cases: Gaussian and Laplace distributions.
- A Lambertian model of reflectance in the scene, which means that the observed light intensity of a given point in the scene is independent from the angle of view, and thus, is equal amongst the views.
- The same color profiles are in the cameras, so that the given light intensity is represented as the same luminance value, Y, among the views (in consideration of a given pair of corresponding pixels $L_{l,y}$ in the left view and $R_{r,y}$ in the right view).

As is has been shown in [29], the assumptions turn out to be true for a variety of multiview video materials. Although some of the assumptions do not hold strictly true for natural sequences (e.g. distribution of noise is only very similar to Gaussian, but does not pass the chi-square test—see Fig. 1, for example), we can use them without loss of conciseness.

3.1.1 Gaussian Probability Distribution of Noise

Let us first consider the presence of Gaussian noise. For such a presence, the conditions mentioned above can be mathematically expressed as

$$L_{l,y} \sim Gaussian_{(Y,\sigma)} \qquad (21)$$
$$R_{r,y} \sim Gaussian_{(Y,\sigma)}$$

where $Gaussian_{(Y,\sigma)}$ is a normal probability distribution with mean value Y and standard deviation σ.

The term $p\left(L_{x+d_{x,y},y} | R_{x,y}, d_{x,y}\right)$ is considered, and thus, random variable $R_{x,y}$ is assumed to be an a posteriori observation with a given, concrete value (also because $d_{x,y}$ is considered conditionally, too), and thus, $Y = R_{x,y}$. Therefore, the pixels are assumed to correspond to each other, and thus, both random variables have the same expected value: $Y_{x,y}$. Moreover, the difference in luminance between $L_{x+d_{x,y},y}$ and $R_{x,y}$ results only from the probability distribution $Gaussian_{(R_{x,y},\sigma)}\left(L_{x+d_{x,y},y}\right)$ for noise, where both $R_{x,y}$ and $L_{x+d_{x,y},y}$ are the a posteriori observations:

$$\mathrm{p}\left(\mathrm{L}_{\mathrm{x+d_{x,y},y}} | \mathrm{R}_{\mathrm{x,y}}, \mathrm{d}_{\mathrm{x,y}}\right) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{\left(\mathrm{L}_{\mathrm{x+d_{x,y},y}} - \mathrm{R}_{\mathrm{x,y}}\right)^2}{2\sigma^2}\right) \qquad (22)$$

and therefore, we get

$$p_{x,y,d} = \frac{p\left(d_{x,y}\right)}{p\left(L_{x,y}\right)} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot exp\left(-\frac{\left(L_{x+d_{x,y},y} - R_{x,y}\right)^2}{2\sigma^2}\right) \qquad (23)$$

We are looking for maximum a posteriori probability, and thus, we search for the best matching disparity d that has the highest (maximal) probability $p_{x,y,d}$. It is equivalent to finding d with maximal $log\left(p_{x,y,d}\right)$ after a natural logarithm on both sides of Eq. (23) is taken:

$$log\left(p_{x,y,d}\right) = log\left(p\left(d_{x,y}\right)\right) - \log\left(p\left(L_{x,y}\right)\right) - log\left(\sigma\sqrt{2\pi}\right) - \frac{\left(L_{x+d_{x,y},y} - R_{x,y}\right)^2}{2\sigma^2} \qquad (24)$$

We can see that if all terms except the last are omitted, Eq. (24) simplifies to a squared differences formula for a pixel similarity metric:

$$log\left(p_{x,y,d}\right) = -\frac{1}{2\sigma^2}\left(L_{x+d,y} - R_{x,y}\right)^2 \qquad (25)$$

The terms omitted in such a way—$\left(d_{x,y}\right)$, $p\left(L_{x,y}\right)$ and $log\left(\sigma\sqrt{2\pi}\right)$—correspond to probability distribution of disparity values, probability distribution of luminance values in the left view, and constant offset, respectively. Such an omission could be justified if all of those terms were constants, which would be true if both of the mentioned probability distributions were uniform.

Of course, if Eq. (25) is applied in blocks of pixels, it corresponds to a sum of squared differences (SSD) similarity metric. We can thus conclude that usage of the SSD metric is optimal (from a maximum a posteriori probability point of view) for cases with the presence of additive Gaussian noise, independence between the views, uniformity in distributions of disparities and luminance values, and with a Lambertian model of reflectance.

3.1.2 Laplace Probability Distribution of Noise

Now, let us consider the presence of Laplace distribution of noise. If such is assumed, similarly (as in the case of Gaussian), we express the following:

$$L_{l,y} \sim Laplace_{(Y,b)} \qquad (26)$$
$$R_{r,y} \sim Laplace_{(Y,b)}$$

where $Laplace_{(Y,b)}$ is the Laplace probability distribution with mean value Y and attenuation parameter b.

Analogous to the case of Gaussian distribution above, we come to the following conclusion if the probability distribution is in form of a Laplace function:

$$p_{x,y,d} = \frac{p\left(d_{x,y}\right)}{p\left(L_{x,y}\right)} \cdot \frac{1}{2 \cdot b} \cdot exp\left(- \frac{\left| L_{x+d_{x,y},y} - R_{x,y} \right|}{b} \right) \qquad (27)$$

and by using the same trick (as with Gaussian noise) of taking a logarithm on both sides of Eq. (27):

$$log\left(p_{x,y,d}\right) = log\left(p\left(d_{x,y}\right)\right) - \log\left(p\left(L_{x,y}\right)\right) - log\left(2 \cdot b\right) - \frac{\left| L_{x+d_{x,y},y} - R_{x,y} \right|}{b} \qquad (28)$$

Here, we can see that if all terms except the last one (on the right) are omitted, Eq. (28) simplifies to an absolute difference formula for a pixel similarity metric:

$$log\left(p_{x,y,d}\right) = -\frac{1}{b}\left| L_{x+d_{x,y},y} - R_{x,y} \right| \qquad (29)$$

Again, the omitted terms $p\left(d_{x,y}\right)$, $p\left(L_{x,y}\right)$, and $log\left(2 \cdot b\right)$ correspond to probability distribution of disparity values, probability distribution of luminance values in the left view, and constant offset, respectively. Such an omission could be justified if all of those terms were constants, which would be true if both of the mentioned probability distributions were uniform.

Similar to the case of SSD, here, we can conclude that usage of the SAD metric is optimal (from a MAP point of view) in the presence of additive Laplace noise, independence between the views, uniformity in distributions of possible disparities and luminance values, and with a Lambertian model of reflectance.

The above-mentioned theoretical derivations are novel, mainly because they show a set of conditions, which if met in a practical case, indicate that usage of SAD or SSD is optimal from a MAP optimization point of view. Of course, the presented reasoning does not limit the application of SAD or SSD pixel similarity metrics to the presented cases only, and thus, usage of SAD or SSD can be found to be optimal in other cases and under optimization on a different basis than MAP.

3.2 The Proposed Probability Model for the *DataCost* Function

The main idea of the proposal is that instead of performing the mentioned simplification to the derived formula in Eq. (16), it can be used directly as formulation for the *DataCost* function.

As a reminder, the formula in Eq. (17) describes a maximum a posteriori probability where, for a given pixel with coordinates x, y, disparity i has occurred:

$$p_{x,y,d} = \frac{p\left(\left(L_{x+d_{x,y},y}, R_{x,y}\right) | v_{x,y,d}\right) \cdot p\left(d_{x,y}\right)}{p\left(L_{x,y}\right) \cdot p\left(R_{x,y}\right)} \qquad (30)$$

In order to use this formula directly, all of the terms of probability in Eq. (30) have to be modeled. Fortunately, all of the required terms have already been measured [29], and those results can be used as follows.

The probability distribution of luminance values in left view $p\left(L_{x,y}\right)$ and in right view $p\left(R_{x,y}\right)$ are calculated as histograms of the input pictures, because those terms do not depend on pixel correspondence related to disparity $d_{x,y}$.

Probability distribution of disparity $p\left(d_{x,y}\right)$ and probability of corresponding luminance values in the left and the right views, $p\left(\left(L_{x+d_{x,y},y}, R_{x,y}\right) | d_{x,y}\right)$, depend on disparity $d_{x,y}$. Having a ground truth disparity map for the given scene, both of those terms can be directly modeled.

- $p\left(d_{x,y}\right)$, which is probability distribution of disparity $d_{x,y}$, has been estimated as a histogram of the given ground truth disparity maps. An example is given in Fig. 2.
- $p\left(\left(L_{x+d_{x,y},y}, R_{x,y}\right) | d_{x,y}\right)$ is a two-dimensional probability distribution that has been estimated as a two-dimensional histogram of luminance values $L_{x+d_{x,y},y}$ and $R_{x,y}$ of pixel pairs, which are known to correspond to each other, based on given disparity value $d_{x,y}$ from the ground truth disparity map. The results attained in [29] have been used (see Fig. 3).

Finally, having all of the terms measured, we can express DataCost for pixel p (with coordinates x, y) to be equal to the expression presented in Eq. (30) on a logarithmic scale. Usage of the logarithmic scale is a

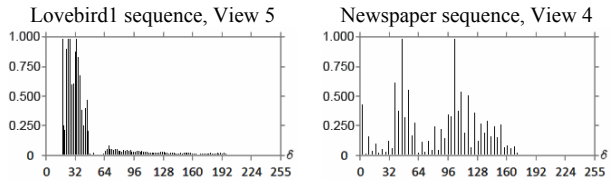


Fig. 2. Examples of histograms of normalized disparity values of pixels in depth maps. The graphs have been normalized to the range [0;1].

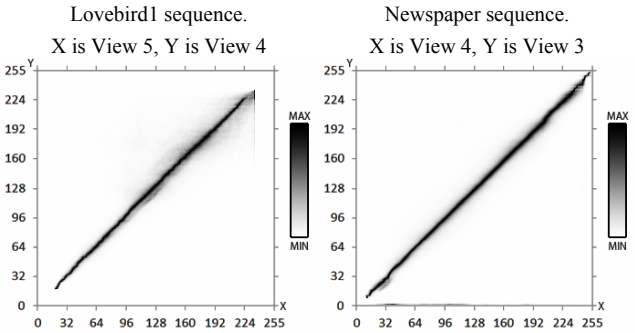


Fig. 3. Two-dimensional histograms of luminance values (on a logarithmic gray-level scale) for corresponding pixels in views X and Y for two multiview test sequences according to measurements from [29].

common trick used in formulation of energy and probability functions for optimization algorithms [2, 4]. We obtain *DataCost* as follows:

$$DataCost_{x,y}\left(d_{x,y}\right) = -10 \cdot log\left(p_{x,y,d}\right) \tag{31}$$

which can be simplified to

$$\begin{aligned} &DataCost_{x,y}\left(d_{x,y}\right) = \\ &10 \cdot \log\left(p\left(\left(L_{x+d_{x,y},y}, R_{x,y}\right) | d_{x,y}\right)\right) - 10 \cdot \log\left(p\left(d_{x,y}\right)\right) \\ &+10 \cdot log\left(p\left(L_{x,y}\right)\right) + 10 \cdot log\left(p\left(R_{x,y}\right)\right). \end{aligned} \tag{32}$$

The final formulation of *DataCost* defined in Eq. (32) is expressed as a logarithmic scale, because the state-of-the-art depth estimation algorithms use it for calculations [10, 11]. Therefore, such a formulation allows for direct application of the proposal, e.g. a graph cuts algorithm implements it in MPEG Depth Estimation Reference Software [25].

4. The Proposed Probability Model for *TransitionCost* Function

As mentioned, in the state-of-the-art depth estimation techniques, the *TransitionCost* function between disparities d_p and d_q of neighboring pixels p and q is denoted as $TransitionCost_{p \to q}\left(d_p, d_q\right)$. In most of the state-of-the-art depth estimation techniques, *TransitionCost* is typically simplified as a function of a single argument: $\left|d_p - d_q\right|$. Examples are the Potts model [16] in Eq. (3), the linear model [3, 25] in Eq. (4), and the truncated-linear model [23] in Eq. (5).

Such usage of those models is arbitrary, for at least two reasons.

- The relation between the probability of disparity between neighboring nodes is typically not measured empirically, and therefore, any assumption about the correctness of a given *TransitionCost* model can be verified only by performing depth estimation.
- All of the mentioned *TransitionCost* models incorporate constant parameters, e.g. α and γ in Eqs. (3)-(5). Those constants are typically chosen experimentally, which is done with limited precision (for example, only four different values of α are tested).

In this paper, a probabilistic model for *TransitionCost* is proposed. Similar to Section 3, a theoretical formulation will be shown, which will then be verified with use of an empirical estimation based on ground truth data.

The proposal employs the assumption that $TransitionCost_{p \to q}\left(d_p, d_q\right)$ can be modeled based on the probability that any two given neighboring pixels p and q have disparities d_p and d_q, respectively. This will be denoted as two-dimensional probability distribution $p_{2D}\left(d_p, d_q\right)$ for the sake of brevity and distinction between pixel p and one-dimensional probability distribution $p_{1D}\left(\cdot\right)$, which will be introduced later.

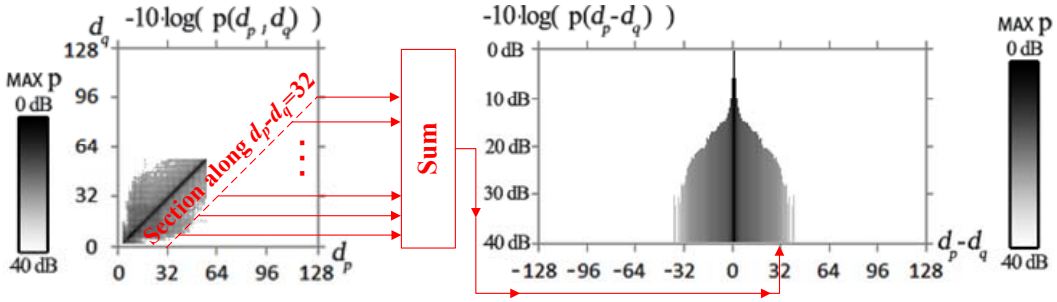
It is assumed that the considered probability distribution $p_{2D}\left(d_p, d_q\right)$ is independent from the position of pixels p and q in the image, and the only constraint is that pixels p and q are direct neighbors.

Therefore, we can express *TransitionCost* on a logarithmic scale so it can be used directly inside state-of-the-art depth estimation algorithms [25]:

$$TransitionCost_{p \to q}\left(d_p, d_q\right) = -10 \cdot log\left(p_{2D}\left(d_p, d_q\right)\right) \tag{33}$$

The main idea of the proposal, as in Section 3, is that instead of making assumptions about the shape of the *TransitionCost* function, it will be measured empirically, based on the ground-truth data available for the test sequences.

The formulation of *TransitionCost* defined in Eq. (33) depends on probability distribution $p_{2D}\left(d_p, d_q\right)$. For real data, it can be measured as a two-dimensional histogram of disparity value pairs d_p and d_q of neighboring pixels p and q. This was performed over all frames of all used test



Distribution of probability $p_{2D}(d_p, d_q)$ where neighboring pixels p and q in the ground truth disparity map, have disparity values d_p and d_q, as the plot of a two-dimensional histogram.

Distribution of probability $p_{1D}(d_p - d_q)$ where neighboring pixels p and q in the ground truth disparity map have difference of disparities $d_p - d_q$, as a plot of a one-dimensional histogram, calculated with Eq. (34). An example of the calculation for $p_{1D}(d_p - d_q = 32)$ is shown in red.

Fig. 4. Probability distributions of disparity values d_p and d_q of neighboring pixels p and q. Both histograms are presented in logarithmic scale and in the same shading, where black reflects the maximum probability value, and white reflects very small probability (-40 dB).

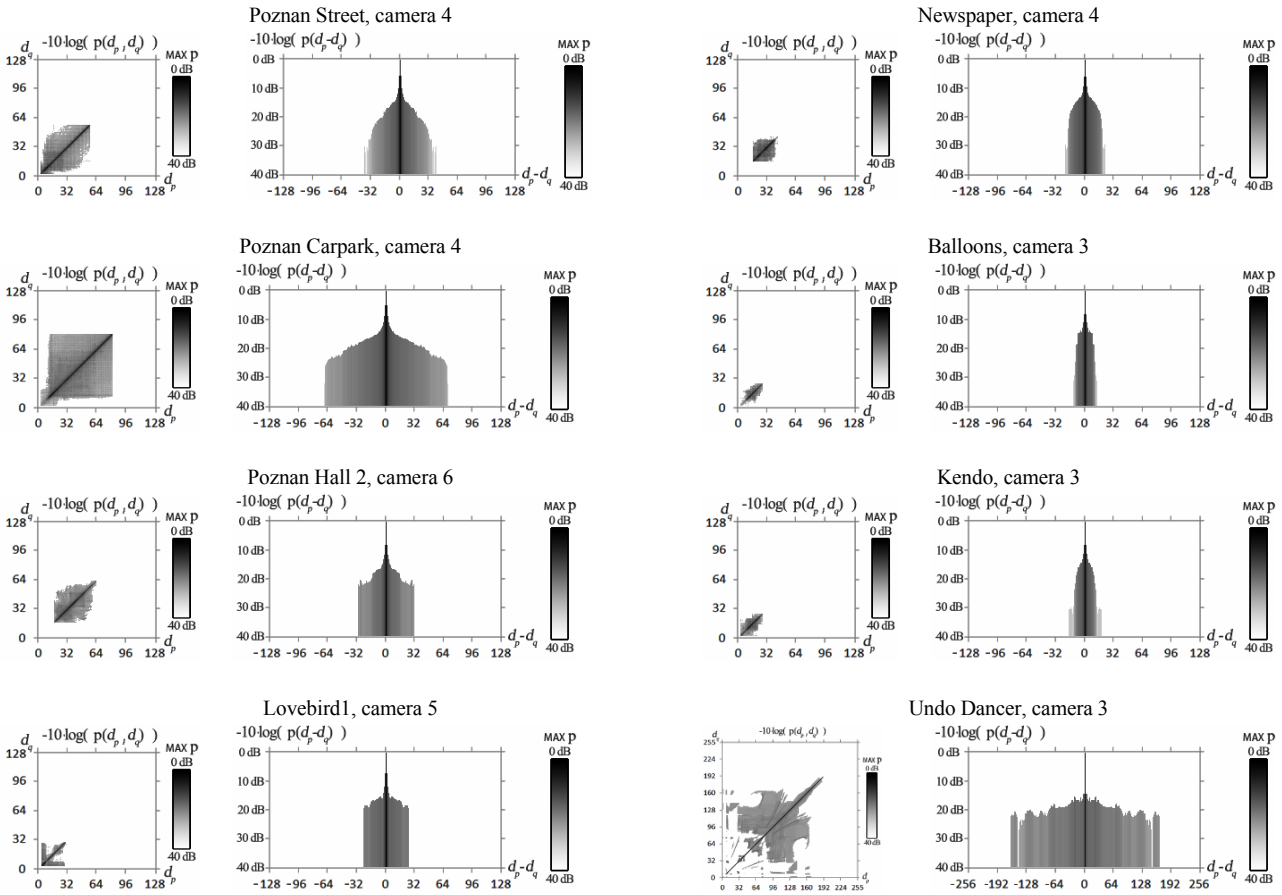


Fig. 5. Histograms of the neighboring disparity values: d_p and d_q - ground truth disparity maps for selected test sequences. The histograms are visualized as 2D plots (left) and histograms in the domain of $(d_p - d_q)$ disparity difference (right). All plots are presented in logarithmic scale and in the same shading. See Fig. 4 for an explanation.

sequences and in all views for which ground-truth depth data are available (see Section 5).

Some of the results (exemplary histograms per sequence) are presented in Figs. 4 (left column) and 5 (left column). We can see that the maximum of the curves lies approximately along the diagonal, but there are also strong bands on both sides. Such strong bands in the histogram means that for given pixel p with disparity d_p, in any neighboring pixel q, a value of disparity d_q is likely to occur if it lies within the probability band of disparity d_p.

Because TransitionCost is often expressed as a function of a single argument, $|d_p - d_q|$, instead of two independent arguments, d_p, d_q—e.g. see Eq. (3) [16] or Eqs. (4) and (5) [3, 23, 25]—it is interesting to also see whether such a formulation is justified. In order to do that, apart from figures presenting $p_{2D}(d_p, d_q)$ as two-dimensional plots (e.g. in Fig. 4 on the left), one-dimensional plots of the $p_{1D}(d_p - d_q)$ probability of given disparity difference $d_p - d_q$ have also been visualized (also see Fig. 5) such that

$$p_{1D}(d_p - d_q) = \sum_{i,j \text{ such that } i-j=d_p-d_q} p_{2D}(i,j) \qquad (34)$$

The results are shown on the right sides of Figs. 4 and 5. Having a look these presented one-dimensional distributions of $d_p - d_q$ (expressed in logarithmic scale), one can notice that the plots first fall approximately linear and then plateau until the limits of the histogram plot. Such plots resemble the shapes of the linear model (Eq. (4)) and the truncated-linear model (Eq. (5)) for *TransitionCost* (see Fig. 6 for comparison).

Therefore, we can conclude that those classical models (linear and truncated-linear) may be adequate for cases when the *TransitionCost* expressed probability is in logarithmic scale (in which *TransitionCost* has been depicted in Figs. 4-6). Important fact is that in each sequence *TransitionCost* has different scale. Without information coming from empirical analysis of *TransitionCost*, executed as in this paper, this scale would have to be calibrated manually or experimentally (e.g. with use of a smoothing coefficient in DERS [25]). This is an important advantage of the proposal presented in this paper.

5. Results

In Sections 3 and 4, probabilistic models were proposed for *DataCost* and *TransitionCost*, respectively. The functional advantages of the proposals were presented, which include lack of the need for manual calibration of parameters.

In this section, an experimental assessment of those models will be provided. Those two proposals together provide a complete model for the *FitCost* function which, as mentioned in Eq. (1), is a sum of *DataCost* and

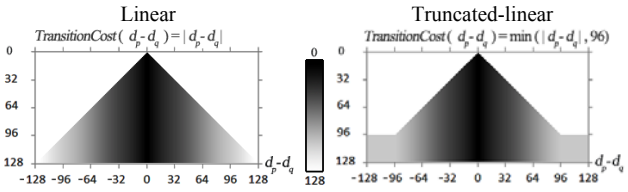


Fig. 6. Examples of graphs for classical TransitionCost functions known from the references: linear (left) and truncated-linear (right). This figure is provided for comparison with graphs in the right columns of Figs. 4 and 5.

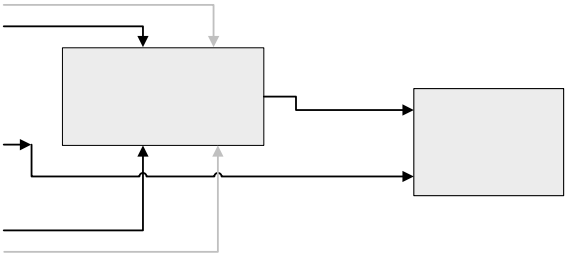


Fig. 7. Depth map assessment procedure developed by ISO/IEC MPEG and used in this paper.

TransitionCost functions. Such a *FitCost* function model will be used in the experimental assessment described below.

The proposed *DataCost* and *TransitionCost* models were implemented in MPEG DERS version 5.1 [25]. The tests were performed following the depth map quality evaluation methodology developed by the ISO/IEC MPEG group, which was constituted as part of the 3D framework [30]. It employs view synthesis for evaluation of the quality of depth maps, which can be used to evaluate the depth estimation algorithm itself.

During the evaluation, three views were explicitly considered: A, B, and V (see Fig. 7). First, for view A and view B, depth maps were estimated. Typically, this is performed with implicit use of some side views. Depth estimation may employ many views (e.g. views A-1, A, and A+1 for depth estimation of view A). The estimated depths of view A and view B, along with their original images, were used to synthesize a virtual view in the position of middle view V.

The original image of view V is used for reference and comparison, which provides indirect evaluation of the depth map estimation algorithm used. Therefore, the quality of the depth was assessed indirectly by evaluation of the quality of the synthesized view. In the methodology developed by ISO/IEC MPEG, for the sake of synthesis of virtual views, usage of View Synthesis Reference Software (VSRS) [21, 32] is recommended. For the purpose of view synthesis, VSRS was also used in this paper.

The used test sequences and view settings are described in Table 1. The model parameters for *DataCost* and *TransitionCost*, estimated with the methods described in Section 3 and Section 4, were used.

The original (unmodified) DERS algorithm is a supervised algorithm in the sense that a special control

Table 1. Test sequences and views selected for evaluation of depth estimation.

Sequence Name	Resolution	Views used for depth estimation (View A and B)	Synthesized view (View V) used for quality evaluation
Poznan Street	1920 × 1088	3, 5	4
Poznan Carpark			
Poznan Hall 2		5, 7	6
Lovebird1	1024 × 768	3, 5	4
Newspaper		4, 6	5
Balloons		3, 5	4
Kendo			

Table 2. Gains obtained with joint usage of the proposed *DataCost* and *TransitionCost* models, related to the best and the worst results obtained with the original (unmodified) DERS, depending on smoothing coefficient parameter settings.

Sequence Name	PSNR [dB] – virtual view versus the original view. Virtual view was synthesized with use of disparity maps with "full-pixel" precision, estimated with the use of:		
	Original (unmodified) DERS: the worst setting of the smoothing coefficient	Original (unmodified) DERS: the best setting of the smoothing coefficient	Proposed probabilistic model implemented in DERS
Poznan Street	27.56	31.98	32.02
Poznan Carpark	29.05	30.71	30.95
Poznan Hall 2	32.17	32.85	32.81
Lovebird1	27.09	29.80	29.83
Newspaper	27.86	31.91	31.95
Balloons	29.95	32.94	32.98
Kendo	33.02	35.46	35.69
Average	29.53	32.24	32.32
Avg. gain with the proposed method related to the references	+2.79	+0.08	-

parameter—the smoothing coefficient—has to be given. Therefore, a wide range of smoothing coefficients was tested. For the sake of brevity, the best and the worst performing settings for each sequence were identified.

The overall results are presented in Table 2. The proposed probabilistic model is similar to the best case of the original (unmodified) DERS in most cases, and is a little better in some cases.

On average over the tested sequences, the proposed method provides about a 0.08 dB gain over the best identified case generated by the original (unmodified) DERS (with a manually crafted smoothing coefficient per sequence) and provides about a 2.79 dB gain over the worst case generated by DERS.

The most important thing to note is that the proposed depth estimation technique does not require any manual settings (usage of such depth estimations is thus unsupervised). The employed *FitCost* function model, based on a maximum a posteriori rule, is inferred with knowledge from analysis of TransitionCost . Therefore, the proposed depth map estimation method was tested only once in one configuration.

6. Conclusions

In this paper, we proposed a complete probabilistic model for the *FitCost* optimization function used in depth estimation, composed of *DataCost* and *TransitionCost* terms. The considerations start with a general theoretical derivation of *DataCost* based on the maximum a posteriori probability rule (see Section 3). On that basis, it was demonstrated that the derived formula can be simplified into classical forms of similarity metrics used in depth estimation, namely, sum of squared differences and sum of absolute differences. Such an observation is interesting because it comes with a set of assumptions that have to be met in order to justify usage of such similarity metrics. Next, a probabilistic model for *TransitionCost* was shown (see Section 4). For both of the proposed models, an empirical method was proposed for estimation of their parameters. Also, the models were used to propose a novel depth estimation method based on a maximum a posteriori probability (MAP) rule. The proposed method allows for unsupervised depth estimation without the usage of arbitrary settings or with manually set control parameters (like a smoothing coefficient in depth estimation reference software) while providing quality in

generated depth maps comparable to cases when supervised depth estimation is used and such parameters are manually optimized. In that case, when sub-optimal settings for control parameters in supervised depth estimation with DERS are used, the proposed method provides a gain of about 2.8 dB measured in average PSNR (Peak Signal-to-Noise Ratio) quality of virtual views synthesized with generated depth maps in the tested sequence set. When optimal settings of control parameters in supervised depth estimation with DERS are used, the gains are negligible, e.g. 0.08 dB, but still they come without the need for manual control parameter selection. Nevertheless, the novelty of the paper is related mostly to general theoretical consideration of the *FitCost* function.

Acknowledgement

The work has been funded by the Polish Ministry of Science and Higher Education within the DS project.

References

R. Jain, R. Kasturi, B. G. Schunck, "Machine Vision", *International Edition, ISBN/ASIN: 0070320187, ISBN-13: 9780070320185, McGraw-Hill*, 1995. Article (CrossRef Link)

M.F. Tappen, W.T. Freeman "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters", *IEEE International Conference on Computer Vision,* 2003. Article (CrossRef Link)

D.M. Greig, B.T. Porteous, A.H. Seheult, "Exact maximum a posteriori estimation for binary images", *Journal of the Royal Statistical Society Series B, 51*, pages 271–279, 1989. Article (CrossRef Link)

Y. Boykov, O. Veksler, R. Zabih, „Markov random fields with efficient approximations", *International Conference on Computer Vision and Pattern Recognition (CVPR),* 1998. Article (CrossRef Link)

[Y. Boykov, O. Veksler, R. Zabih, „Fast approximate energy minimisation via graph cuts", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 29, pages 1222-1239, 2001. Article (CrossRef Link)

L.R. Ford Jr., D.R Fulkerson, "Maximal flow through a network", *Canadian Journal of Mathematics,* Vol. 8, pages 399-404, 1956. Article (CrossRef Link)

P. Elias, A. Feinstein, C.E. Shannon "Note on maximum flow through a network", *IRE Trans. on Information Theory, IT*. 2, No. 4, pages 117-119, 1956. Article (CrossRef Link)

Y. Boykov, V. Kolmogorov "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision", *IEEE Transactions on Pattern Anal. Mach. Intell.* 26(9): 1124-1137, 2004. Article (CrossRef Link)

Christos H. Papadimitriou, Kenneth Steiglitz, „The max-flow, min-cut theorem", chapter in "Combinatorial Optimization: Algorithms and Complexity", *2nd edition, Dover,* pages 120–128. ISBN 0-486-40258-4, 1998. Article (CrossRef Link)

[D. Scharstein, R. Szeliski, "Middlebury Stereo Vision Page", Article (CrossRef Link)
online 1st Dec 2013. Article (CrossRef Link)

D. Scharstein, R. Szeliski, „A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *International Journal of Computer Vision 2002*. Article (CrossRef Link)

M.I. Jordan "Learning in Graphical Models", *MIT Press, ISBN: 9780262600323,* January 1999. Article (CrossRef Link)

[I.J. Cox, S.L. Hingorani, S.B. Rao, B.M. Maggs "A maximum-likelihood stereo algorithm", *Computer Vision and Image Understanding,* Vol. 63, No. 3, pages 542-567, 1996. Article (CrossRef Link)

J. Sun, N.-Ning Zheng, H.-Y. Shum, „Stereo matching using belief propagation", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 25, Issue: 7, pages 787 - 800, 2003. Article (CrossRef Link)

L. Cheng, T. Caelli "Bayesian stereo matching", *Proc. Conf. Computer Vision and Pattern Recognition Workshop*, pages 192-192, 2004. Article (CrossRef Link)

S. Geman, G. Geman "Stochastic Relaxation, Gibbs distribuition and the Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 6, pages 721-741, 1984. Article (CrossRef Link)

S.Z. Li. "Markov Random Field Modeling in Image Analysis", *ISBN: 978-1-84800-278-4, Springer,* 2009. Article (CrossRef Link)

O. Stankiewicz, K. Wegner, M. Domański, „Depth estimation based on maximization of A posteriori probability" International Conference on Computer Vision and Graphics ICCVG 2016, Warsaw, Poland, 19-21 September 2016. Article (CrossRef Link)

T. Kanade, M. Okutomi "A stereo matching algorithm with an adaptive window: theory and experiment", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 16, No. 9, pages 920-932, Sept. 1994. Article (CrossRef Link)

Y. Boykov, O. Veksler, R. Zabih "A variable window approach to early vision", *IEEE Trans. Pattern Anal. Mach. Intell.,* Vol. 20, No. 12, pages 1283–1294, Dec. 1998. Article (CrossRef Link)

R. Zabih, J. Woodfill, "Non-parametric local transforms for computing visual correspondence", *Proceedings of European Conference on Computer Vision,* Springer, Berlin, Heidelberg, 1994. Article (CrossRef Link)

K. Wegner, O. Stankiewicz, „Similiarity measures for depth estimation", *3DTV-Conference 2009 The True Vision Capture, Transmission and Display of 3D Video,* Potsdam, Germany, 4-6 May 2009. Article (CrossRef Link)

Li Zhang, S.M. Seitz, „Estimating optimal parameters for mrf stereo from a single image pair", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol.:29 , Issue: 2, pages 331 - 342, 2007. Article (CrossRef Link)

P. Viola, W. Wells, "Alignment by maximization of mutual information", *Proceedings of Fifth International Conference on Computer Vision,* pages 16-23, Cambridge, MA, USA, 20-23 Jun 1995. Article (CrossRef Link)

M. Tanimoto, T. Fujii, K. Suzuki, „Video Depth Estimation Reference Software (DERS) with image segmentation and block matching", *ISO/IEC JTC1/SC29/WG11* Doc., M16092, Lausanne, Switzerland, Feb. 2009. Article (CrossRef Link)

K. Wegner, O. Stankiewicz M. Domański, „Stereoscopic depth estimation using fuzzy segment matching", *28th Picture Coding Symposium (PCS2010),* Nagoya, Japan, 8-10 Dec. 2010. Article (CrossRef Link)

Z. Zhang, Determining the epipolar geometry and its uncertainty: A review", *International Journal of Computer Vision, 27(2):*161-1195, 1998. Article (CrossRef Link)

C. Doutre, P. Nasiopoulos, "Color correction preprocessing for multiview video coding,", *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 19, No. 9, pages 1400-1405, Sep. 2009. Article (CrossRef Link)

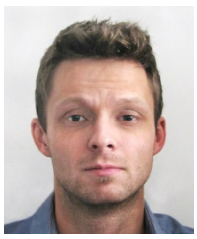
O. Stankiewicz, M. Domański, K. Wegner, „Analysis of noise in multi-camera systems", 3DTV Conference 2014, Budapest, Hungary, 2-4 July 2014. Article (CrossRef Link)

"Overview of 3D video coding", *ISO/IEC JTC1/SC29/WG11,* Doc.N9784, Archamps, France, May 2008. Article (CrossRef Link)

M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, Y. Mori, "Reference softwares for depth estimation and view synthesis", *ISO/IEC JTC1/SC29/WG11,* Archamps, France, Tech. Rep. M15377, Apr. 2008. Article (CrossRef Link)

"View synthesis algorithm in view synthesis reference software 3.0 (VSRS3.0)", *ISO/IEC JTC1/SC29/WG11* Doc. M16090, Feb.2009. Article (CrossRef Link)

Marek Domański received an MSc, a PhD, and a Habilitation degree from Poznań University of Technology, Poland, in 1978, 1983, and 1990, respectively. Since 1993, he has been a Professor at Poznań University of Technology, Chair (Department) of Multimedia Telecommunications and Microelectronics. Since 2005, he has been head of the Polish delegation to MPEG. He co-authored highly ranked technology proposals submitted in response to MPEG calls for scalable video compression (2004) and 3D video coding (2011). He also led the team that developed one of the first AVC decoders for TV set-top boxes (2004) and various AVC, HEVC, and AAC HE codec implementations and improvements. He is the author of three books and more than 300 papers for journals and proceedings of international conferences. The contributions were mostly on imaging, video and audio compression, virtual navigation, free viewpoint television, image processing, multimedia systems, 3D video and color image technology, digital filters, and multidimensional signal processing. He is a co-inventor in several patents and pending patent applications in European and US patent offices. He was General Chairman/Co-Chairman and host of several international conferences: the Picture Coding Symposium, PCS 2012, IEEE Int. Conf. on Advanced and Signal-based Surveillance, AVSS 2013, the European Signal Processing Conference, EUSIPCO 2007, the 73rd and 112th Meetings of the MPEG Int. Workshop on Signals, Systems and Image Processing, IWSSIP 1997 and 2004 Int. Conf. of Signals and Electronic Systems, and ICSES 2004, among others. He served as a member of various steering, program, and editorial committees for international journals and international conferences.



Olgierd Stankiewicz received his PhD from the Faculty of Electronics and Telecommunications, Poznań University of Technology, in 2014. Currently, he is an Assistant Professor for Chair of Multimedia Telecommunications and Microelectronics, where he has been working since receiving his Master of Engineering degree in 2006. In 2005, he won second place in the IEEE Computer Society International Design Competition (CSIDC) held in Washington, D.C. He is actively involved in ISO standardization activities, where he contributes to the development of 3D video coding standards. From 2011 to 2014, he was a coordinator for the development of MPEG reference software for 3D video coding standards based on AVC. He actively contributed to JCT-3V, MPEG-I, MPEG Free Viewpoint TV, and JPEG-PLENO exploration and standardization activities. He has published more than 100 papers (journals, proceedings of international conferences, and MPEG/JPEG databases) on free viewpoint television, depth estimation, view synthesis, and hardware implementations in FPGAs. His professional interests include signal processing, video compression algorithms, computer graphics, and hardware solutions. He is a co-inventor in several patents and pending patent applications in European and US patent offices.