Volume 0 (1981), Number 0 pp. 1–12 COMPUTER GRAPHICS *forum*

Application-Specific Tone Mapping via Genetic Programming

K. Debattista

WMG, University of Warwick, UK
K.Debattista@warwick.ac.uk

Abstract

High dynamic range (HDR) imagery permits the manipulation of real-world data distinct from the limitations of the traditional, low dynamic range (LDR), content. The process of retargetting HDR content to traditional LDR imagery via tone mapping operators (TMOs) is useful for visualising HDR content on traditional displays, supporting backwards-compatible HDR compression and, more recently, is being frequently used for input into a wide variety of computer vision applications. This work presents the automatic generation of TMOs for specific applications via the evolutionary computing method of genetic programming (GP). A straightforward, generic GP method that generates TMOs for a given fitness function and HDR content is presented. Its efficacy is demonstrated in the context of three applications: visualisation of HDR content on LDR displays, feature mapping and compression. For these applications, results show good performance for the generated TMOs when compared to traditional methods. Furthermore, they demonstrate that the method is generalisable and could be used across various applications that require TMOs but for which dedicated successful TMOs have not yet been discovered.

Categories and Subject Descriptors (according to ACM CCS): I.4.0 [Image Processing and Computer Vision]: General—Image processing software

1. Introduction

High dynamic range (HDR) imaging permits the capture, manipulation and display of real-world lighting [BADC11]. HDR content differs from its low/standard dynamic range (LDR) equivalent due to its ability to store real-world values represented as floating point data as opposed to the single byte per channel approach of traditional imagery. While HDR displays are becoming more common the majority of displays do not support HDR and this will not change in the immediate future [CD17]. In such cases the HDR data needs to be converted into content suitable for LDR displays. This problem is generally tackled by a group of techniques called tone mapping operators (TMOs) which convert the HDR signal into an LDR one. Furthermore, recent years have seen a rise in the use of HDR data for computer vision applications [PCZ*16] to deal with harsher lighting conditions, which may require the use of software that is LDR specific, hence a conversion from HDR to LDR is still required.

A large number of TMOs do already exist but they are typically generic and made primarily for users to visualise HDR content on LDR displays. Certain applications could benefit from tone mapped data that is designed explicitly for that application. For example, Mai et al. [MMM*11] designed a TMO which was optimised for compression of HDR content. Yet, this is one of the few cases of such designs. Designing TMOs for particular applications may be time consuming and the costs may outweigh the benefits if significant research is required for a particular application. In this work the process of identifying an application-specific tone mapper suitable for a task is automated by the use of genetic programming (GP) [Koz92]. GP is a method that generates programs (or functions) by adapting them over generations in an evolutionary computing environment based on a fitness function corresponding to the problem being solved.

This work makes use of GP for solving the tone mapping problem across various domains. The main contribution of this work is to show that GP can be practical for application-specific problems that require TMOs. A generic tone mapping GP model is developed and is used to create new TMOs for applications. Three general applications are shown in order to demonstrate the potential of this approach: traditional tone mapping for visaulisation of HDR content on LDR displays, tone mapping for feature detection and tone mapping for backwards-compatible HDR compression. Results, presented for each of the methods are comparable with state-of-the-art methods and can be used for similar problems. Fol-

lowing this GP framework, it should be possible to generate new TMOs for problems not presented here, or with different goals than those shown in these applications.

2. Background and Related Work

This section presents related work in tone mapping for various applications, provides some background on GPs and related work in that domain.

2.1. Tone Mapping

There is a significant amount of literature on TMOs particularly in the case of generic TMOs developed for visualising HDR content on traditional displays [BADC11]. Broadly, the methods could be divided into two categories global and local TMOs. The global TMOs apply a function to the entire image, typically based on global characteristics of the image, while local TMOs modify individual pixels or neighbourhoods based on local statistics. Global operators include a number of linear and logarithmic methods, brightness reproduction [TR93], the Schlick method [Sch94] and the adaptive logarithmic TMO of Drago et al. [DMAC03]. The local operators include the spatial non-uniform scaling [CHS*93] and the photographic tone reproduction operator (PTRO) [RSSF02]. Mantiuk and Siedel [MS08] present an alternative TMO method that can generically model TMOs and successfully replicates many traditional TMOs. Mantiuk and Siedel's method is related to the work presented here as it explores a broad search space of TMOs, however, it is more specific to replicating traditional methods while the GP method proposed is a framework for generating new TMOs for diverse applications. Ma et al. [MYZW15] presented a framework for optimising TMOs based on the metric TMQI-II presented in the same work.

A number of recent publications have explored and evaluated TMOs in feature mapping applications. They have, generally, adopted a similar approach, consisting of the evaluation of key-point detection across various TMO methods. Rana et al. [RVD15, RVD16] present the use of key-point detection using both SURF and Harris detectors. Přibyl et al. [PCZ*16] also presented a similar method with further TMOs tested and other detection methods. Kontogianni et al. [KSGD15] analysed the performance of TMOs on SIFT, SURF, and ORB. Suma et al. [SSS*16] produced a similar evaluation with the focus on architectural reconstructions. Furthermore, the use of TMOs has been used directly within computer vision applications such as facial recognition [PMPP14, KBPE15], facial expression recognition [IDC16] and tracking [ASGD15].

TMOs are also used for HDR compression for backwards-compatible HDR compression methods which maintain two streams, a tone mapped stream that is viewable on traditional displays and a residual stream. In the Ward and Simmons' [WS04] and Mantiuk et al.'s [MEMS06] HDR compression methods the TMO that generates the backwards compatible stream is user defined; however the performance of different TMOs on the resultant compression is important and was evaluated for both methods. Mai et al. [MMM*11] designed a TMO specifically for optimising HDR compression. They presented a framework whereby a piecewise linear tone curve was optimised for better compression results and performed better than traditional TMOs.

2.2. Genetic Programming

GP [Koz92] is an evolutionary computing method closely related and developed from genetic algorithms that represents the genotype as computer programs or functions instead of the traditional encoding strings used in genetic algorithms. Such methods provide solutions to problems using the natural selection concept of evolving generations, whereby offspring are generated from the healthiest individuals in the environment. The healthiest individuals are selected based on a fitness function which evaluates the performance of the individual at solving the problem proposed. GP permits the exploration of a significant number of solutions to a given problem by generating and evolving programs for the task at hand based on a defined grammar that dictates program construction. In a given generation individual programs are tested for their performance and the healthiest ones are used to generate the next population via a series of evolutionary inspired operators such as crossover and mutation.

GPs have frequently been used for image classification and pattern matching problems; Espejo et al. [EVH10] present a comprehensive overview. They have also been frequently used for generating features too, for example for facial expression recognition [YB06] or epileptic EEG classification [GRD*11]. GPs have also been used for image compression [NB96] and video compression [KLTN02]. Within computer graphics there have been a few applications of GPs. Kensler and Shirley [KS06] used them to generate faster ray triangle intersection operators for use in ray tracing computations. Sitthi-Amorn et al. [SAMWL11] used GP for the automatic simplification of procedural shaders. Similarly, Brady et al. [BLPW14] made use of GP to generate new analytical BRDFs which compared favourably with state-of-the-art methods. Kán et al. [KDK17] presented a GP for finding new Monte Carlo noise filters.

3. Method

The proposed GP method is intended to automatically generate TMOs that can solve problems such as those outlined in Section 2.1. TMOs are typically based on a function that takes as input real-world valued data to produce display adapted LDR data:

$$L_d = f(L_w), \text{ where } f : \mathbb{R}^+ \rightarrow [0, 255] \quad (1)$$

$$
\begin{array}{lll}
< node > & ::= & < operation > | < terminal > \\
< operation > & ::= & < unaryOp > | < binaryOp > \\
< unaryOp > & ::= & < unaryFn > (< node >) \\
< binaryOp > & ::= & < node >< binaryFn >< node > \\
< unaryFn > & ::= & \log \mid \log_2 \mid \log_{10} \mid \sin \mid \cos \mid \tan \mid \text{asin} \mid \text{acos} \mid \text{atan} \mid \exp \mid \text{sqrt} \\
< binaryFn > & ::= & + \mid - \mid \times \mid / \mid \hat{} \\
< terminal > & ::= & L_w(x) \mid \text{<constant>} \mid \text{<global>} \\
< constant > & ::= & \text{rand}(-5 \ldots 5) \mid 1 \mid 2 \mid 3 \\
< global > & ::= & L_{\max} \mid L_{\text{mean}} \mid L_{\text{median}} \mid L_{\text{harMean}} \mid L_{\text{mean}(\log_{10})} \mid L_{\exp(\text{mean}(\log))}
\end{array}
$$

Figure 1: *Application-specific TMO grammar.*

where L_w corresponds to the HDR content, L_d to the tone mapped LDR results and $f(.)$ is the TMO. For global operators $f(.)$ is a function applied equally to all content in the original HDR image. The goal of the GP presented in this section is to identify a suitable TMO $f(.)$ for a given application.

Algorithm 1 Genetic Programming overview

1: $generation \leftarrow 0$
2: $p \leftarrow$ generateInitialPopulation()
3: **repeat**
4: $\quad f \leftarrow$ getFitness(p)
5: $\quad p \leftarrow$ generateNextPopulation(p, f)
6: $\quad generation \leftarrow generation + 1$;
7: **until** $generation = maxGenerations$

3.1. Overview

GP follows a general structure, shown in Algorithm 1. A population is initially created (line 2), its fitness is computed (line 4), individuals are selected based on their fitness and a new population is generated from the previous based on the selected individuals (line 5). This cycle is run across a given number of generations (line 7).

In this work a multi-island model in which each island runs the cycle from Algorithm 1 independently on subpopulations and migrates individuals at given intervals is used. In a multi-island model each island progresses in parallel thus developing semi-independent populations concurrently permitting faster computation through computational parallelism. Such semi-independent populations permit autonomous development, potentially reducing the stagnation of similar high fitness individuals that could lead to being stuck in local optima, while providing cross-fertilisation via migration [CHMR87]. A star island migration method was employed as it was found to produce good results compared to other migration methods [AGC12]. The star island model migrates the best individuals to a given island at chosen intervals.

The goal of this GP is to discover an application-specific function $f(.)$ from Equation 1. $f(.)$ is allowed to be generated via the GP based on the grammar presented in Figure 1. Frequently used arithmetic and geometric terms are used. The $< globals >$ correspond to commonly used global computations in TMOs, many of them can be generated automatically but some are popular enough to warrant separate inclusion. $L_{\max}$, L_{mean}, L_{median}, L_{harMean}correspond to the maximum, mean, median and harmonic mean of the input image. $L_{\text{mean}(\log_{10})}$is computed as $\frac{1}{N}\sum \log_{10}(x)$ and $L_{\exp(\text{mean}(\log))}$as $\exp(\frac{1}{N}\sum \log(x))$.

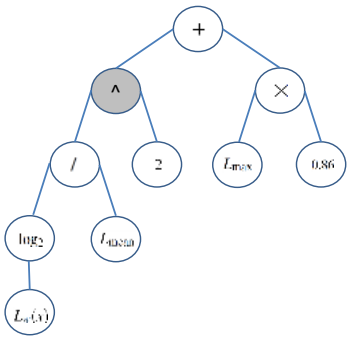
3.2. Population Creation

Populations of individuals are initially created randomly without any code blocks, from existing TMOs or data from previous runs. This is done to maintain the simplicity of the method and has provided good results. The method employed for creation is based on the grow method [PLMK08] whereby trees are allowed to be grown randomly and terminate at any depth up to a given fixed depth at which a terminal must be chosen.

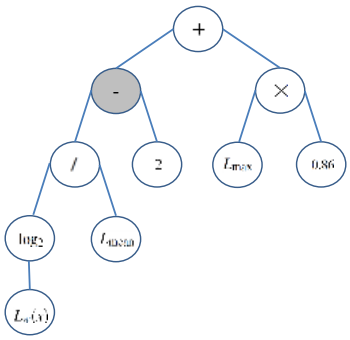
3.3. Fitness Function and Selection

The fitness function describes the performance of a given individual for a particular task. In this case it calculates the performance of the generated TMO on a particular set of HDR images. This section presents the overall GP method that can be applied to identify TMOs for specific applications. The main application specific aspect depends on an application-dependent fitness function; the different fitness functions for particular applications will be discussed in Section 4. The rest of the GP method assumes that this function is given and nothing else needs to be modified.

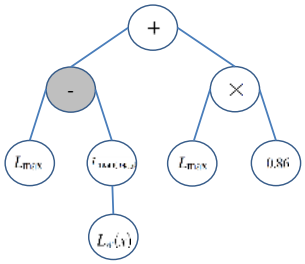
The choice of fitness is based on fitness proportionate representation. In this method individuals are chosen for generating the next generation based on their proportionate fitness compared to the rest of the population. This is calculated by first constructing the normalised cumulative distribution function (CDF) of the fitness results for the entire population, choosing a random number between 0 and 1, and identifying the individual chosen based on the CDF.



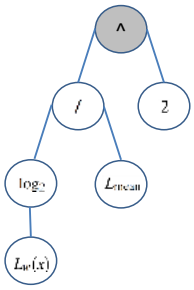
(a) Original: $\left(\frac{\log_2(L_w(x))}{L_{\text{mean}}}\right)^2 + (0.86 \times L_{\max})$



(b) NR: $\left(\left(\frac{\log_2(L_w(x))}{L_{\text{mean}}}\right) - 2\right) + (0.86 \times L_{\max})$



(c) ST: $L_{\max} - \text{ExpLogMean}(L_w(x)) + (0.86 \times L_{\max})$



(d) Hoist: $\left(\frac{\log_2(L_w(x))}{L_{\text{mean}}}\right)^2$

Figure 2: *An illustration of the different mutation strategies. The grey node in (a) is the one chosen to be mutated. (b) (Node Replacement mutation - NR) shows how the $<$ binaryOp $>$ ^ is replaced by $-$ but no other nodes are affected (note the equation can now be simplified but is left in its entirety to be more legible). (c) (Subtree mutation - ST) shows how the entire subtree is substituted (again this can be simplified but is maintained intact). (d) (Hoist mutation) shows how the subtree now constitutes the entire function.*

When evaluating the function a number of considerations are made. All pixels in the image are run through the same function. Initially, the HDR content has a small delta added to it to avoid values of 0. The resulting data is tidied up due to the possible complexities of the operations. Once the function is run, the data is clamped at $[0,1)$, the imaginary component of complex data is discarded and finally the image is gamma corrected and quantised to $[0,255]$ to conform with traditional LDR imagery.

3.4. Generating the Next Population

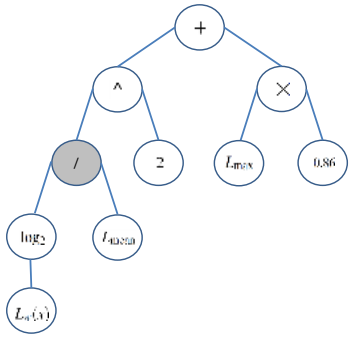
This section outlines the specific operations used to generate the next population from the previous. Details of the proportions of a new population generated by any one of these operations are given in Section 4.1.1.

Elite: a selection of the top individuals is chosen and copied directly to the new generation. Unlike mutation and crossover, the choice of the elite individuals is not based on fitness proportionate selection but on the overall fitness.

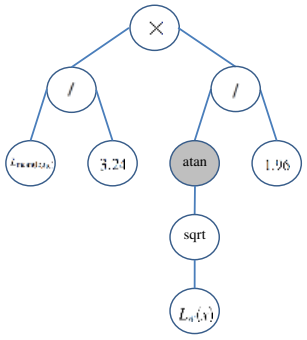
Mutation: Mutation uses fitness proportionate selection to choose an individual. Once the individual is chosen it is mutated. Mutation is based on a mutation rate tested as the nodes are traversed. Three forms of traditional mutation methods are adopted, see Figure 2:

Node replacement mutation: A node in the individual's subtree is chosen at random. The element in the node is replaced by another element of the same type in the grammar, e.g., a $<$ *binaryFn* $>$ such as '+' is replaced with another $<$ *binaryFn* $>$ '$\times$. This is permissable for nodes of type $<$ *unaryFn* $>$, $<$ *binaryFn* $>$, $<$ *constant* $>$ and

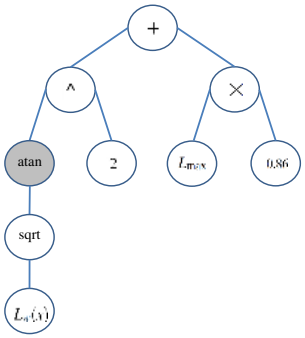
$< global >$. All elements in the subtree of the node remain unchanged. This mutation permits small but potentially meaningful changes to the fitness of an individual.



(a) First individual: $\left(\frac{\log_2(L_w(x))}{L_{\text{mean}}}\right)^2 + (0.86 \times L_{\max})$



(b) Second individual: $\frac{L_{\text{mean}(\log_{10})}}{3.24} \times \frac{\text{atan}(\sqrt{L_w(x)})}{1.96}$



(c) Result: $\text{atan}(\sqrt{L_w(x)})^2 + (0.86 \times L_{\max})$

Figure 3: *An illustration of cross over. The grey node in (a) and (b) are chosen for cross over. The subtree in (b) is crossed over to the subtree in (a) resulting in a new individual (c).*

Subtree mutation: A node is chosen at random and the entire subtree with this node as its root is replaced with a new subtree, created following the same rules based on population creation.

Hoist mutation: A subtree of the individual becomes the root. This mutation helps reduce possible bloat if aspects of an individual's function are not contributing much to the computation. This function is equivalent to a crossover of an individual's root with one of its own subtrees.

If the mutation fails, i.e., no random node was chosen, a new individual is created from scratch to permit the addition of new individuals to the population.

Crossover: For crossover two individuals are chosen by fitness proportionate selection and used to generate a single new individual. The subtree of one individual is then replaced by the subtree of the other, see Figure 3. The choice of which subtrees crossover depends on the crossover rate, analogous to the mutation rate, that selects subtrees from the two chosen individuals. Failed crossovers, that did not chose subtrees are replaced by new individuals.

A simplification function is also run once every generation to ensure that subtrees that do not contain $< globals >$ or $L_w(x)$ are resolved immediately.

4. Applications

This section introduces three applications to which the method is applied to demonstrate the potential of identifying distinct solutions adapted for various problems. These are traditional tone mapping, feature matching with key point detection and HDR image compression. The section commences with some implementation details and settings used across the three applications.

4.1. Applying GPs for Application Specific Tone Mapping

For each application two broad types of runs can be executed, and this is how these applications are presented. A generic run that is exposed to a number of images (or image pairs) and produces a general function for that particular application and a second individual run that produces functions for each image (or image pair) seen. For the generic run only a small subset of the content used in the application is used in the GP run. The individual run effectively results in a TMO particularly optimised for the given content. In an ideal situation all content is run through its own individual GP run; however, this comes at an extra computational cost which the generic run alleviates.

4.1.1. Overall Settings

The overall settings remain the same across the generic and individual runs and across all three presented applications. The only two things that are different for the applications

are the loading of data and the fitness function calculation. In terms of the implementation these are passed as function pointers to the GP. Data loading is run upon initialisation and the fitness function is run once for each generation. If a TMO for a new application needed to be generated it would suffice to create these two functions. For the generic and individual runs the only difference is in the data presented and the the total number of generations.

The sub-populations on each island consist of 250 individuals. In the population creation phase the depth of the function trees are limited to 10. After each generation the top 2% of individuals are considered elite and are maintained and the rest of the 98% are split evenly between mutation and crossover based on the fitness proportionate selection method as discussed previously. The mutation rate was set to 0.25 per node for all mutations. Similarly, the crossover rate was set to 0.25.

As mentioned above, a multi-island model using star island migration is adopted. Four islands are run over 100 generations for the generic case and 25 for the individual case. These numbers are chosen to compromise between performance and the ability to traverse and refine a significant search space. Migrations occur every five generations, following Brady et al. [BLPW14]. The three best methods from each island were migrated to the first island during these migration occurrences. The first island ran an additional five generations after the final migration for a total of 105 migrations for the generic and 30 for the individual runs. The winner was the best method in terms of fitness to be located on the first island.

4.1.2. Implementation, Comparisons and Evaluations

The method was implemented and tested in Matlab using functionality from the HDR toolbox [BADC11]. The winning GP TMOs are compared with a number of TMOs. The Photographic Tone Reproduction Operator [RSSF02], in both its local (PTRO) and global (PTRO-global) incarnations was chosen as this method frequently performs well against other methods in subjective evaluations [LCTS05]. The Drago [DMAC03] (Drago), Shclick [Sch94] (Schlick) and Fattal [FLW02] (Fattal) methods are chosen as they have been frequently evaluated in feature mapping methods; the Drago TMO has been known to perform well in some studies [RVD15, SSS*16], and the Fattal TMO well for others [PCZ*16]. The HDR Toolbox [BADC11] implementations of the above TMOs are employed. The optimal method [DBRS*15] (optimal) is also used as it is conceptually similar to the automatic exposure performed by cameras, and produces good results for feature matching and compression. Finally, the compression-specific tone mapper by Mai et al. [MMM*11] (Mai) is employed for the compression application. This has been implemented in Matlab for the purposes of comparison and implements the optimisation aspect of the method only, without the distortion component; however, this should be sufficient to provide an overview of the best performance when JPEG is not used (this is not used for the proposed GP method either when generating results). Details of the application-specific evaluation method used are outlined in the individual application sections below. Default values were maintained for all methods used.

4.2. Traditional Tone Mapping

In this section the performance of the method to generate a traditional TMO used for visualisation of HDR content on LDR displays is tested.

4.2.1. Overview

A large number of TMOs exist that perform the general tone mapping task of converting an HDR image for viewing on a traditional display device. There have been a significant number of studies that compared across them [BADC11]. While this work was not designed to challenge such methods, as further validation of the system a TMO resulting from a GP run is compared with traditional TMOs.

4.2.2. GP Method

The fitness function adopted for this test was TMQI [YW13] a metric designed for comparing images generated by TMOs with the original HDR content, although other metrics such as TMQI-II [MYZW15] could also have bene used. Such metrics make the evaluation of such methods possible. The data used during the GP run were six images downscaled to half their original resolution. These images were obtained from the subset of 36 images taken from the Fairchild database used in perceptual studies of dynamic range [HDVD16]. This database of images were chosen because of their availability and because they had been selected as they broadly cover the HDR range in the afore-mentioned study. The six images used in the GP generic run are shown in Figure 4. The winning run was:

$$\arctan(\arctan\left(\frac{0.1411 \times L_w(x)}{L_{\exp(\text{mean}(\log))}}\right)). \tag{2}$$

The individual run was also conducted for each the six images in Figure 4.

4.2.3. Results and Evaluation

To ensure running the method is practical, results comparing the proposed method with state-of-the-art methods are based on the full set of 36 HDR images at HD resolution used in the perceptual study [HDVD16]. They are further augmented for this test by the 19 left eye images from the set of images used for the feature mapping application discussed below and the Stanford Memorial Chapel image for a total of 56 images.

Table 1 presents the results of the method. The proposed method outperforms all other methods in this test. The local

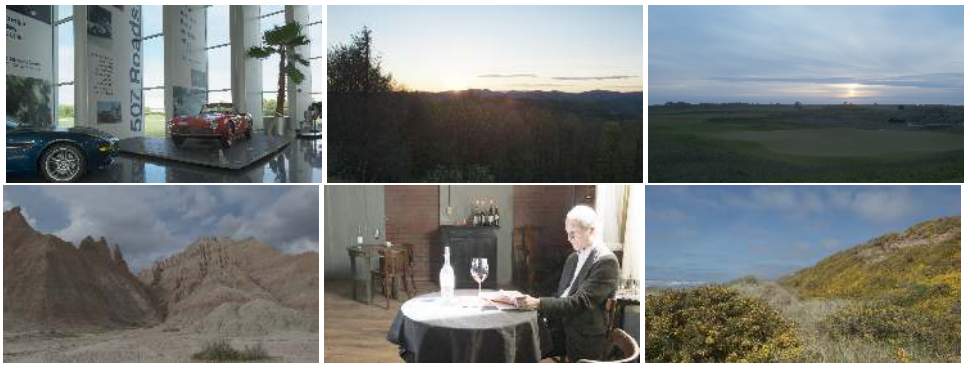


Figure 4: *The images used in the traditional tone mapping and compression GP runs. Images tone mapped using Equation 2.*

Method	TMQI score
Prop. - Eqn. 2	0.8641
PTRO (local)	0.8560
PTRO (global)	0.8493
Schlick	0.7081
Drago	0.8373
Fattal	0.7939
Optimal	0.8250
†Prop. - Eqn. 2	0.8302
†Individual	0.8775
‡Prop. - Eqn. 2	0.8679

Table 1: *TMO comparison for humans using the TMQI metric for comparing. Mean results across 56 images are presented.* † *results for six images in Figure 4 only.* ‡ *results for the remaining 50 images not used in the GP run.*

PTRO operator, which frequently performs well in subjective evaluations of tone mappers, follows closely in second place. The global version follows with the other methods following behind. The table also shows results for images not used in the GP run and results can be broadly seen to be the same. The results for the individual run are averaged over the six images and demonstrate an improvement over the winning method for the six images tested. Figure 6 shows results of different TMOs against those generated by the GP runs for two images.

4.3. Featuring Mapping

As the ability to generate HDR images becomes more widespread the use of HDR for computer vision applications is bound to increase. Many such applications, for example those built in hardware, may expect LDR images as input, hence it is not uncommon to use tone mappers to convert the HDR images for input into such systems. In the following the GP is applied to feature matching, as it is one of the base steps of many vision methods, following the overall methodology described in recent publications [KSGD15, RVD16, SSS*16].

4.3.1. Overview

The method for evaluating feature mapping follows the evaluation of certain key-point detectors across traditional TMOs [PCZ13,RVD15]. In particular Rana et al. [RVD15] is followed in the choice of the two key-point detection methods chosen, however, any other method could have been adopted and used for the fitness function and for evaluation. The first feature mapping method tested is the Harris corner and edge detector [HS88] and the second the SURF blob detector method [BTVG06].

Both feature matching methods are evaluated based on two metrics. The first is the straightforward metric consisting of the number of features detected $|F|$ across a pair of images. While this is useful it is not always ideal, hence other metrics such as the repeatability rate (RR) are frequently employed as an alternative measure [SMB00]. RR is computed as $\frac{|F|}{min(n_i,n_k)}$ where n_i and n_k are the number of points detected in images I_i and I_k and $|F|$ is the number of features matched as above. RR has also been adopted for similar tests in previous evaluation studies [RVD15, SSS*16]. Figure 7 shows the matching for two superimposed image pairs from the set used for the GP run in Figure 5.

4.3.2. GP Method

Four GPs were run with four distinct fitness functions corresponding to the two feature mapping methods (Harris, SURF) and the two metrics ($|F|$, RR). The distinction between the Harris and SURF runs was based on only the use of the detection function without and further differences between the two methods. If any other detector were used it would suffice to just change this part of the fitness function.

For the test that optimised for $|F|$ it was not sufficient to merely use the actual number of features as the images produced by the method could be corrupted beyond human

Figure 5: *The images used in the GP run for feature mapping. The left image for each stereo pair is shown.*

(a) Eqn. 2 (b) Individual (c) PTRO (local) (d) Drago (e) Optimal

(f) Eqn. 2 (g) Individual (h) PTRO (global) (i) Fattal (j) Schlick

Figure 6: *Comparison of various TMO outputs.*

comprehension, hence TMQI was also adopted to ensure human viewable images. This was weighted by five times in the fitness calculation compared to the TMQI error based on empirical evidence from pre-study runs. For RR, straight adoption of the metric would favour TMOs that would result in very small numbers of features detected. To avoid this, the fitness function for RR included the $|F|$ and TMQI weighting as in the $|F|$ only case, with RR weighted by 1.5 for similar reasons discussed above.

The fitness function was run on five pairs of stereo HDR images, see Figure 5. These were chosen from the overall stereo HDR data set of nineteen stereo HDR pairs used by Selmanovic et al. [SDBRC12]. The first five pairs in the data set where arbitrarily chosen for GP fitness computation; however, evaluation was conducted on the entire set of nineteen (see below).

The four winning TMOs are presented below. The proposed TMO for Harris feature matching RR was:

$$L_d(x) = \arctan(3.249034 \times L_w(x)). \quad (3)$$

The proposed TMO for Harris using $|F|$ was:

$$L_d(x) = \frac{L_w(x)}{L_{\text{mean}}}. \quad (4)$$

The discovered method for RR for SURF was:

$$L_d(x) = L_w(x)^2 + (1.968137 \times L_w(x)) + L_{\exp(\text{mean}(\log))}. \quad (5)$$

Finally, the method identified for producing the largest number of SURF features $|F|$ was:

$$L_d(x) = \frac{L_w(x)}{e^{L_{\text{mean}(\log_{10})}}}. \quad (6)$$

For the individual run all four GP runs were run separately for each of the image pairs shown in Figure 5.

4.3.3. Evaluation and Results

For evaluation purposes the results are presented to evaluate the four GP runs and resulting TMOs with traditional TMOs. These tests are run over the entire nineteen pairs and results are presented from the full set.

The results for Harris for RR and $|F|$ are presented in Table 2. Similarly, the results for SURF are presented in Table 3 (note that the SURF results include results for Equation 7 introduced below). As shown all the proposed methods perform very well for the metric they were generated for and outperform all other methods except for Fattal outperforming Equation 5 for RR in the SURF results. It is worth noting that the use of weighting $|F|$ for the RR metric across both Harris and Surf means that the proposed RR methods Equation 3 and Equation 5 produce reasonably large $|F|$, the Harris one is only second after the dedicated method and the SURF one third. For SURF, the winning Fattal method has a relatively low $|F|$ which contributes to the high RR. Another GP RR run with a double weighting for RR compared to $|F|$ was run for SURF with all other settings kept the same. The resultant method:

$$L_d(x) = L_w(x) \times (\log_{10}(\log(L_{\max} - 4.115346))), \quad (7)$$

achieves an RR of 0.6245 and $|F|$ of 209.06 outperforming

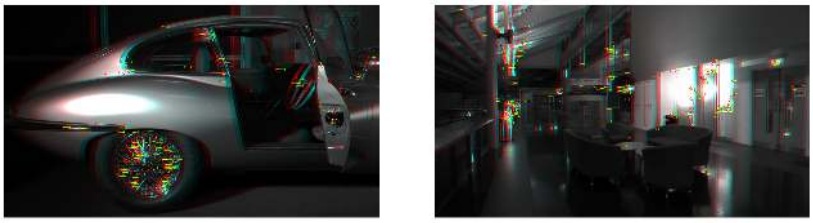


Figure 7: *Feature matching using keypoint detection via SURF.*

Method	RR	$\|F\|$
Prop. - Eqn. 3	0.3798	322.26
Prop. - Eqn. 4	0.3453	353.74
PTRO (local)	0.3327	168.73
PTRO (global)	0.3579	255.11
Schlick	0.3268	236.16
Drago	0.3664	236.15
Fattal	0.3725	200.37
Optimal	0.3757	145.21
†Prop. - Eqn. 3	0.3925	114.40
†Prop. - Eqn. 4	0.3226	152.80
Individual RR	0.4654	81.40
Individual $\|F\|$	0.3088	238.60
‡Prop. - Eqn. 3	0.3753	396.50
‡Prop. - Eqn. 4	0.3542	428.64

Table 2: *TMO comparison for feature matching using Harris. Mean results across 19 image pairs are shown.* † *results for six image pairs shown in Figure 5 only.* ‡ *results for the remaining 14 image pairs never seen during the GP run.*

Method	RR	$\|F\|$
Prop. - Eqn. 5	0.5877	414.00
Prop. - Eqn. 6	0.5606	624.32
Prop. - Eqn. 7	0.6245	209.06
PTRO (local)	0.5667	396.58
PTRO (global)	0.5684	390.74
Schlick	0.5569	491.00
Drago	0.5660	386.00
Fattal	0.5988	209.95
Optimal	0.5855	365.11
†Prop. - Eqn. 5	0.5552	287.80
†Prop. - Eqn. 6	0.5070	400.80
†Prop. - Eqn. 7	0.5645	163.80
Individual RR	0.6889	147.00
Individual $\|F\|$	0.4816	912.20
‡Prop. - Eqn. 5	0.5993	459.07
‡Prop. - Eqn. 6	0.5800	704.35
‡Prop. - Eqn. 7	0.6459	225.21

Table 3: *TMO comparison for feature matching using SURF. Mean results across 19 image pairs are shown.* † *results for six image pairs shown in Figure 5 only.* ‡ *results for the remaining 14 image pairs never seen during the GP run.*

all other methods for RR, but as with Fattal has a low $|F|$. The results of the individual runs are again positive and, as expected, outperform the more generic methods for the four metrics tested.

4.4. Compression

Due to the size of HDR data representations, compression of HDR content remains an important application in HDR research. In order to encourage adoption, several methods use a backwards compatible approach whereby two streams of the content are maintained, one for viewing on standard displays/viewers and a second stream to augment the first for displaying HDR content directly. The first stream makes use of TMOs to generate the content; however, such tone curves are not generated for compression efficiency with the exception of the work of Mai et al. [MMM*11], and even in this work the optimisation was conducted per scenario. In this application GP is used to generate a TMO which optimises for the quality of the compressed content.

4.4.1. Overview

The Ward and Simmons method [WS04] of HDR JPEG compression is used for the compression example presented here because it is straightforward, permits the use of any TMO and has recently been adopted as one of the main profiles for the industry-wide HDR standard JPEG-XT [AMR*15]. Furthermore, other backwards compatible HDR video methods follow similar concepts for example the rate distortion method [LK08] and HDR MPEG [MEMS06]. In the Ward and Simmons method a viewable tone mapped image is stored together with a ratio image (RI) used to reconstruct the original HDR. The HDR image is initially tone mapped with the TMO of choice and a ratio image is generated. The tone mapped image is further processed and derived from the ratio image and original HDR and subsequently both the RI and tone mapped image are compressed using traditional LDR compression methods, such as JPEG.

4.4.2. GP Method

The calculation of the fitness function was based on the methodology followed by Mukherjee et al. [MDBR*16]. In this test the evaluation the TMO and ratio images were first computed using the generated function and then quantized to 8-bit values. They were then de-quantized and the HDR image was reconstructed. The original HDR image was then compared with the reconstructed one using the PU-PSNR metric [AMS08] which correlated very highly with subjective data in HDR compression comparisons [MDBR*16]. The GP method was generated using the same six images used for traditional tone mapping shown in Figure 4. The resultant winner from this run was:

$$L_d(x) = \sqrt{L_w(x) \times \sin(\arctan(\frac{0.846}{L_{\max}}))}. \quad (8)$$

As with the traditional tone mapping an individual run which optimises for the compression of given content is conducted. Six GP runs were also conducted for each of the images in Figure 4 resulting in six distinct functions.

4.4.3. Evaluation and Results

As with the other applications, the evaluation compares the winning method against other commonly used TMOs. The comparison is based on the 56 full sized HDR images used for the first application in order to verify real-world adaptability. The method is tested against the other methods, firstly without any JPEG compression (NC), effectively evaluating the performance of the tone curve as was conducted for the fitness training. Furthermore, results are shown for JPEG compression at quality levels 100, 75, 50 and 25 were 100 is the highest quality and 25 is the lowest quality. Around 75 corresponds to a general default value. For the JPEG tests the ratio image was resized to halve the size and then resized back when expanded.

Results, are shown in Table 4. For the no JPEG compression (NC) case a large improvement for Equation 8 is seen compared to the generic TMOs. This is in part expected as the method was designed to optimise for this value albeit over a much smaller sample. The results for JPEG compression show a similar trend with the proposed function outperforming all others except in the case of Optimal for the quality settings of 100. The Mai method performs well too as expected. The proposed method also produced relatively small sized images for each of the JPEG cases with average values of 1,212KB, 190KB, 126KB and 84KB for the quality settings of 100, 75, 50 and 25 respectively. The average size across all methods were 1,346KB, 226KB, 150KB, 100KB for quality settings of 100, 75, 50, 25 respectively . Note, this method was not designed for this but it is reported to show that improved performance did not necessarily arise at the cost of increase file sizes.

Results for individual runs once again demonstrate an improvement in the results for the images tested.

Method	PU-PSNR score				
	NC	100	75	50	25
Prop. - Eqn. 8	71.34	55.11	50.99	49.13	46.95
PTRO (local)	59.07	52.88	49.79	48.21	46.11
PTRO (global)	62.40	54.25	49.36	47.67	46.29
Schlick	53.53	48.29	46.36	45.08	43.46
Drago	63.47	52.65	49.53	47.99	46.13
Fattal	68.01	51.60	49.15	47.78	46.00
Optimal	62.28	55.89	50.46	48.73	46.52
Mai	69.44	54.08	50.38	48.52	46.16
† Prop. - Eqn. 8	68.95	56.01	51.64	49.72	47.24
† Individual	71.98	57.01	51.67	49.75	47.44
‡ Prop. - Eqn. 8	71.56	55.14	51.04	49.18	47.01

Table 4: *TMO comparison for compression using the PU-PSNR metric. Mean results across 56 images are shown. NC stands for no JPEG compression, the numbers 100 . . . 25 indicate the level of JPEG compression used.* † *results for six images in Figure 4 only.* ‡ *results for the remaining 50 images not used in the GP run.*

5. Discussion, Limitations and Conclusions

GP methods and their implementations can be relatively complex with the number of parameters and design choices required to be made being significant. In this work, purposefully, a straightforward GP method was used such that it is generalisable across a number of applications. In this respect, this work set out to demonstrate the viability of applying GP to the tone mapping problem. It demonstrated that GP is a useful method for generating TMOs that have the ability to perform well for given problems. This characteristic makes this method suitable for those applications where a detailed manual investigation into tone mapping solutions may be unjustified or can be used as a starting point for further investigation. The applications for which it has been used performed well compared to the state of the art, however it is not claimed that, for example, the TMO developed from Equation 2 should be used instead of the already established standard methods - it has just performed better under this given metric. The main contribution of this work lies in its flexibility to adapt to different applications and produce results for applications which have not seen significant contributions in research such as methods for computer vision. Furthermore, it enables per-image (or general content) TMO solutions for a given proposed application. Finally, compared to other optimisation methods the resultant function is instantly legible.

The work also sheds some light on what the general structure of TMOs for certain applications may look like. Equation 2 is not that dissimilar to other general tone mappers as the arctan(arctan()) nature resembles sigmoids used in many tone mappers. The broadly linear nature of many of the winning feature mapping methods also broadly reflects general observations in the field [CD17]. However, further work is

required to better quantify general trends in specific applications.

5.1. Future Work

The results of the individual runs demonstrate the potential of the method for content-specific runs, generating the best solutions for given applications for particular content. Furthermore, these are also more desirable as global runs may provide a solution that is incompatible with data it was not trained on. While attractive, this comes at the cost of performance, and individual runs are not much slower than a full run. Future work will investigate optimisations to the methods to try and identify whether a reasonable cost-benefit performance for individual runs can be achieved.

Currently, only global functions are produced, however, grammars could be adapted to account for local operations also. Local operations could be added via the inclusion a variable to $<terminal>$ in the grammar, see Figure 1, that operates on a local area. This could be expressed as a convolution, not dissimilar to those used in convolutional neural networks [LBBH98]. The convolutions themselves can be modified across generations using evolutionary methods that modify kernel weights, size and strides, using the same operations for generating the next generation as those presented in Section 3.4. $< local >$ could also be added to the grammar to provide local detail similar to what the current $< global >$ accomplishes but within a local neighbourhood. While the current method's straightforward approach was a conscious design decision, it remains to be seen whether more complex designs such as the one described above yield better results for certain applications.

One current trend is the use of temporal TMOs [EMU17] this further increase in complexity becomes a challenge for GPs due to the amount of data required, in order to run the GPs, potentially entire animations need to be passed to the GP and for large populations this can be prohibitive. Furthermore, there are few efficient evaluation metrics for dynamic TMOs rendering the process more complicated. Hopefully, this work will encourage further exploration of the topic that could render GPs for dynamic TMOs a useful research endeavor.

Acknowledgements

Debattista is partially supported by a Royal Society Industrial Fellowship (IF130053). Many thanks to Tom Bashford-Rogers for fruitful discussions on the topic.

References

[AGC12] ANDALON-GARCIA I. R., CHAVOYA A.: Performance comparison of three topologies of the island model of a parallel genetic algorithm implementation on a cluster platform. In *Electrical Communications and Computers (CONIELECOMP), 2012 22nd International Conference on* (2012), IEEE, pp. 1–6. 3

[AMR*15] ARTUSI A., MANTIUK R., RICHTER T., HANHART P., KORSHUNOV P., AGOSTINELLI M., TEN A., EBRAHIMI T.: Overview and evaluation of the JPEG-XT HDR image compression standard. *Journal of Real-Time Image Processing* (2015), 1–16. 9

[AMS08] AYDIN T., MANTIUK R., SEIDEL H.-P.: Extending quality metrics to full luminance range images. In *Electronic Imaging 2008* (2008), International Society for Optics and Photonics, pp. 68060B–68060B. 10

[ASGD15] AGRAFIOTIS P., STATHOPOULOU E. K., GEORGOPOULOS A., DOULAMIS A. D.: HDR Imaging for Enchancing People Detection and Tracking in Indoor Environments. In *VISAPP (2)* (2015), pp. 623–630. 2

[BADC11] BANTERLE F., ARTUSI A., DEBATTISTA K., CHALMERS A.: *Advanced high dynamic range imaging: theory and practice*. CRC press, 2011. 1, 2, 6

[BLPW14] BRADY A., LAWRENCE J., PEERS P., WEIMER W.: genbrdf: Discovering new analytic brdfs with genetic programming. *ACM Transactions on Graphics (TOG) 33*, 4 (2014), 114. 2, 6

[BTVG06] BAY H., TUYTELAARS T., VAN GOOL L.: Surf: Speeded up robust features. In *European conference on computer vision* (2006), Springer, pp. 404–417. 7

[CD17] CHALMERS A., DEBATTISTA K.: HDR video past, present and future: A perspective. *Elsevier Image Communication 54*, October (May 2017), 49–55. http://dx.doi.org/10.1016/j.image.2017.02.003. 1, 10

[CHMR87] COHOON J. P., HEGDE S. U., MARTIN W. N., RICHARDS D.: Punctuated equilibria: a parallel genetic algorithm. In *Genetic algorithms and their applications: proceedings of the second International Conference on Genetic Algorithms: July 28-31, 1987 at the Massachusetts Institute of Technology, Cambridge, MA* (1987), Hillsdale, NJ: L. Erlhaum Associates, 1987. 3

[CHS*93] CHIU K., HERF M., SHIRLEY P., SWAMY S., WANG C., ZIMMERMAN K.: Spatially nonuniform scaling functions for high contrast images. In *Proceedings of Graphics Interface 93* (San Francisco, CA, USA, May 1993), Morgan Kaufmann Publishers Inc., pp. 245–253. 2

[DBRS*15] DEBATTISTA K., BASHFORD-ROGERS T., SELMANOVIĆ E., MUKHERJEE R., CHALMERS A.: Optimal exposure compression for high dynamic range content. *The Visual Computer 31*, 6-8 (2015), 1089–1099. 6

[DMAC03] DRAGO F., MYSZKOWSKI K., ANNEN T., CHIBA N.: Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum 22*, 3 (2003), 419–426. 2, 6

[EMU17] EILERTSEN G., MANTIUK R., UNGER J.: A comparative review of tone-mapping algorithms for high dynamic range video. *STAR 36*, 2 (2017). 11

[EVH10] ESPEJO P. G., VENTURA S., HERRERA F.: A survey on the application of genetic programming to classification. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews 40*, 2 (2010), 121–144. 2

[FLW02] FATTAL R., LISCHINSKI D., WERMAN M.: Gradient domain high dynamic range compression. In *ACM Transactions on Graphics (TOG)* (2002), vol. 21, ACM, pp. 249–256. 6

[GRD*11] GUO L., RIVERO D., DORADO J., MUNTEANU C. R., PAZOS A.: Automatic feature extraction using genetic programming: An application to epileptic eeg classification. *Expert Systems with Applications 38*, 8 (2011), 10425–10436. 2

[HDVD16] HULUSIĆ V., DEBATTISTA K., VALENZISE G., DUFAUX F.: A model of perceived dynamic range for hdr images. *Signal Processing: Image Communication* (2016). 6

[HS88] HARRIS C., STEPHENS M.: A combined corner and edge detector. In *Alvey vision conference* (1988), vol. 15, Citeseer, pp. 10–5244. 7

[IDC16] IGE E. O., DEBATTISTA K., CHALMERS A.: Towards hdr based facial expression recognition under complex lighting. In *Proceedings of the 33rd Computer Graphics International* (2016), ACM, pp. 49–52. 2

[KBPE15] KORSHUNOV P., BERNARDO M. V., PINHEIRO A. M., EBRAHIMI T.: Impact of tone-mapping algorithms on subjective and objective face recognition in HDR images. In *Proceedings of the Fourth International Workshop on Crowdsourcing for Multimedia* (2015), ACM, pp. 39–44. 2

[KDK17] KĂAŅ P., DAVLETALIYEV M., KAUFMANN H.: Discovering New Monte Carlo Noise Filters with Genetic Programming. In *EG 2017 - Short Papers* (2017), Peytavie A., Bosch C., (Eds.), The Eurographics Association. 2

[KLTN02] KRANTZ T., LINDBERG O., THORBURN G., NORDIN P.: Programmatic compression of natural video. *GECCO Late Breaking Papers 2* (2002), 301–307. 2

[Koz92] KOZA J. R.: *Genetic programming: on the programming of computers by means of natural selection*, vol. 1. MIT press, 1992. 1, 2

[KS06] KENSLER A., SHIRLEY P.: Optimizing ray-triangle intersection via automated search. In *Interactive Ray Tracing 2006, IEEE Symposium on* (2006), IEEE, pp. 33–38. 2

[KSGD15] KONTOGIANNI G., STATHOPOULOU E., GEORGOPOULOS A., DOULAMIS A.: Hdr imaging for feature detection on detailed architectural scenes. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 40*, 5 (2015), 325. 2, 7

[LBBH98] LECUN Y., BOTTOU L., BENGIO Y., HAFFNER P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE 86*, 11 (1998), 2278–2324. 11

[LCTS05] LEDDA P., CHALMERS A., TROSCIANKO T., SEETZEN H.: Evaluation of tone mapping operators using a high dynamic range display. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers* (New York, NY, USA, 2005), ACM, pp. 640–648. 6

[LK08] LEE C., KIM C.-S.: Rate-distortion optimized compression of high dynamic range videos. In *16th European Signal Processing Conference (EUSIPCO 2008)* (2008), pp. 461–464. 9

[MDBR*16] MUKHERJEE R., DEBATTISTA K., BASHFORD-ROGERS T., VANGORP P., MANTIUK R. K., BESSA M., WATERFIELD B., CHALMERS A.: Objective and subjective evaluation of high dynamic range video compression. *Signal Processing: Image Communication 47* (2016), 426–437. 10

[MEMS06] MANTIUK R., EFREMOV A., MYSZKOWSKI K., SEIDEL H.-P.: Backward compatible high dynamic range mpeg video compression. *ACM Trans. Graph. 25*, 3 (2006), 713–723. 2, 9

[MMM*11] MAI Z., MANSOUR H., MANTIUK R., NASIOPOULOS P., WARD R., HEIDRICH W.: Optimizing a tone curve for backward-compatible high dynamic range image and video compression. *IEEE transactions on image processing 20*, 6 (2011), 1558–1571. 1, 2, 6, 9

[MS08] MANTIUK R., SEIDEL H.-P.: Modeling a generic tone-mapping operator. *Computer Graphics Forum 27*, 2 (April 2008), 699–708. 2

[MYZW15] MA K., YEGANEH H., ZENG K., WANG Z.: High dynamic range image compression by optimizing tone mapped image quality index. *IEEE Transactions on Image Processing 24*, 10 (2015), 3086–3097. 2, 6

[NB96] NORDIN P., BANZHAF W.: Programmatic compression of images and sound. In *Proceedings of the 1st annual conference on genetic programming* (1996), MIT Press, pp. 345–350. 2

[PCZ13] PŘIBYL B., CHALMERS A., ZEMČÍK P.: Feature point detection under extreme lighting conditions. In *Proceedings of the 28th Spring Conference on Computer Graphics* (2013), ACM, pp. 143–150. 7

[PCZ*16] PŘIBYL B., CHALMERS A., ZEMČÍK P., HOOBERMAN L., ČADÍK M.: Evaluation of feature point detection in high dynamic range imagery. *Journal of Visual Communication and Image Representation 38* (2016), 141–160. 1, 2, 6

[PLMK08] POLI R., LANGDON W. B., MCPHEE N. F., KOZA J. R.: *A field guide to genetic programming*. Lulu. com, 2008. 3

[PMPP14] PEREIRA M., MORENO J.-C., PROENÇA H., PINHEIRO A. M.: Automatic face recognition in hdr imaging. In *SPIE Photonics Europe* (2014), International Society for Optics and Photonics, pp. 913804–913804. 2

[RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic tone reproduction for digital images. *ACM Trans. Graph. 21*, 3 (2002), 267–276. 2, 6

[RVD15] RANA A., VALENZISE G., DUFAUX F.: Evaluation of feature detection in hdr based imaging under changes in illumination conditions. In *Multimedia (ISM), 2015 IEEE International Symposium on* (2015), IEEE, pp. 289–294. 2, 6, 7

[RVD16] RANA A., VALENZISE G., DUFAUX F.: Optimizing tone mapping operators for keypoint detection under illumination changes. In *2016 IEEE Workshop on Multimedia Signal Processing (MMSP 2016)* (2016). 2, 7

[SAMWL11] SITTHI-AMORN P., MODLY N., WEIMER W., LAWRENCE J.: Genetic programming for shader simplification. *ACM Transactions on Graphics (TOG) 30*, 6 (2011), 152. 2

[Sch94] SCHLICK C.: Quantization techniques for visualization of high dynamic range pictures. In *Proceeding of the Fifth Eurographics Workshop on Rendering* (June 1994), pp. 7–18. 2, 6

[SDBRC12] SELMANOVIC E., DEBATTISTA K., BASHFORD-ROGERS T., CHALMERS A.: Backwards compatible jpeg stereoscopic high dynamic range imaging. In *TPCG* (2012), pp. 1–8. 8

[SMB00] SCHMID C., MOHR R., BAUCKHAGE C.: Evaluation of interest point detectors. *International Journal of computer vision 37*, 2 (2000), 151–172. 7

[SSS*16] SUMA R., STAVROPOULOU G., STATHOPOULOU E. K., VAN GOOL L., GEORGOPOULOS A., CHALMERS A.: Evaluation of the effectiveness of hdr tone-mapping operators for photogrammetric applications. *Virtual Archaeology Review 7*, 15 (2016), 54–66. 2, 6, 7

[TR93] TUMBLIN J., RUSHMEIER H.: Tone reproduction for realistic images. *IEEE Comput. Graph. Appl. 13*, 6 (1993), 42–48. 2

[WS04] WARD G., SIMMONS M.: Subband encoding of high dynamic range imagery. In *APGV '04: Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization* (New York, NY, USA, 2004), ACM Press, pp. 83–90. 2, 9

[YB06] YU J., BHANU B.: Evolutionary feature synthesis for facial expression recognition. *Pattern Recognition Letters 27*, 11 (2006), 1289–1298. 2

[YW13] YEGANEH H., WANG Z.: Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing 22*, 2 (2013), 657–667. 6