Computer Vision and Image Understanding 152 (2016) 67–78

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

A mutual local-ternary-pattern based method for aligning differently exposed images

Shiqian Wu[a,*], Lingxian Yang[b], Wangming Xu[b], Jinghong Zheng[c], Zhengguo Li[c], Zhijun Fang[d]

[a] School of Machinery and Automation, Wuhan University of Science and Technology, 430081, China
[b] School of Information Science and Engineering, Wuhan University of Science and Technology, 430081, China
[c] Department of Signal Processing, Institute for Infocomm Research, 1 Fusionopolis Way, 21-01 Connexis, 138632, Singapore
[d] School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, 333 Long Teng Rd, Shanghai, 201620, China

ARTICLE INFO

Article history:
Received 17 November 2015
Revised 1 May 2016
Accepted 29 July 2016
Available online 30 July 2016

Keywords:
Image registration
Local ternary pattern
Multi-exposed images
Image mapping function
Histogram-based matching
Coarse-to-fine technique
HDR imaging

ABSTRACT

Saturation and large intensity variations occurred in multi-exposed images offer great challenges to align these images. In this paper, a mutual local-ternary-pattern (MLTP) is proposed to represent differently exposed images for image registration. Different from the classical local ternary pattern (LTP) and its variants, the proposed MLTP has two salient properties: (1) The ternary pattern of one image is not only determined by itself, but also relied on its counterpart; (2) The MLTP is grayscale-adaptive. It is analyzed that the proposed MLTP is a good representation to preserve consistency of differently exposed images. Based on the MLTP-coded images, an efficient linear model derived from Taylor expansion is presented to estimate motion parameters. To improve accuracy and efficiency, image rotation is initially detected by the histogram-based matching, and coarse-to-fine technique is implemented to cope with possibly large movement. Extensive experiments carried out on a variety of synthesized and real multi-exposed images demonstrate that the proposed method is robust to 10 exposure values (EV), which is superior to other methods and current commercial HDR tools.

1. Introduction

In the current world where storage, bandwidth, processing power and sophisticated devices have provided an excellent convenient infrastructure for people to capture, utilize and consume images, there has been an increasing demand for high quality in the manipulated images as compared to what human eyes can see. As a result, super resolution, extended depth of field, high dynamic range (HDR) imaging and other image enhancements by photographic multi-shot techniques have been used in many applications (Juergen and Rainer, 2009). As the input of these technologies is a sequence of images, image registration is crucial to obtain high-quality synthesized output. Currently, combining differently exposed low dynamic range (LDR) images of the same scene has been the most popular approach to generate an HDR image. This paper focuses on aligning multi-exposed images captured by hand-held devices for HDR imaging, which is still a challenging issue in image registration (Oldridge et al., 2011).

1.1. Alignment of multi-exposed images – the state of the arts

Image registration or alignment has been a fundamental problem in image processing and computer vision. There are a large number of techniques proposed for a variety of applications, for example, spatial variations, intensity variations, sensor variations and so on. The papers (Oldridge et al., 2011; Szeliski, 2006) provide a comprehensive survey on image registration. Generally, the registration methods can be classified into two categories (Szeliski, 2006): pixel-based (area-based) methods and feature-based methods. Pixel-based methods find motion parameters through minimizing pixel-to-pixel dissimilarities. Accordingly, such methods are dependent on image intensities. Feature-based methods, on the other hand, first extract distinctive features from each image, e.g., point features (Lowe, 2004), line features (Maintz et al., 1996) etc., then match and warp the features to derive parametric transformations. As feature-based methods do not work directly with image intensities, it is frequently used when illumination (intensity) changes.

So far, feature-based techniques have been adopted to align a set of differently exposed images. In Eden et al. (2006); Tomaszewska and Mantiuk (2007), SIFT (Scale-Invariant Feature

* Corresponding author. Fax: +862768862283.
E-mail address: shiqian.wu@wust.edu.cn (S. Wu).

http://dx.doi.org/10.1016/j.cviu.2016.07.010
1077-3142/

Transform) method was employed to detect feature points (keypoints) in multi-exposed images, and the RANSAC method was used to find best pairs of keypoints and derive motion parameters. In Gevrekci and Gunturk (2007), Gevrekci and Gunturk proposed to detect corners as feature points. To alleviate the influence of intensity variations on extracting feature points, both the algorithms (Gevrekci and Gunturk, 2007; Tomaszewska and Mantiuk, 2007) work in contrast domain, which is further revealed in Gevrekci and Gunturk (2009) that the repeatability rate of the detected features can be improved by about 25%. Meanwhile, pixel-based methods were employed for this task, in which the key idea is to cope with intensity variations. In Cerman and Hlavac (2006); Kang et al. (2003), the multi-exposed images are modeled via camera response function (CRF) proposed by Debevec and Malik (1997). Then, the differently exposed images can be converted to identical exposure via the CRF. Another technique is to perform joint geometric and photometric registration (Aguiar, 2006; Bartoli, 2008; Candocia, 2003; 2005; Hossain and Gunturk, 2011; Luong et al., 2010; Zimmer et al., 2011). In Aguiar (2006); Luong et al. (2010), the intensities of multi-exposed images are modeled as a linear relationship. A two-step iterative algorithm in Aguiar (2006) and total least square in Luong et al. (2010) were proposed to solve geometric registration. In Bartoli (2008), Bartoli modeled photometric mapping as gain and offset terms, and estimated the motion field as well as the photometric mapping parameters. Based on gradient constant assumption, gradient information was used for alignment in Zimmer et al. (2011). In Hossain and Gunturk (2011), two differently exposed images were normalized by the intensity mapping function. In Candocia (2003), Candocia proposed to model multi-exposed images by a nonlinear model which was introduced in Mann (2000), spatial and tonal registrations were then simultaneously performed by Levenberg-Marquardt optimization. The key drawback of this method is that the optimization is very slow as there are many parameters, i.e., $9(q-1)$ parameters, where q is the number of images, to be estimated simultaneously. Moreover, optimization in high dimension is difficult to guarantee global solution. To mitigate the computation burden, an improved solution using piecewise linear comparametric model was proposed in Candocia (2005). A hybrid scheme employing a pixel-based method as well as a feature-based method was proposed in Tico and Pulli (2010), and a two-stage method comprising of image normalization and local-binary-pattern representation was developed in Wu et al. (2014) for registering multi-exposed images. Recently, patch-based methods have been prevalent (Hu et al., 2012; 2013; Ramirez et al., 2013; Sen et al., 2012; Zheng and Li, 2015; Zheng et al., 2013) in HDR imaging. The key advantage of these approaches lies in simultaneously dealing with camera movement and object movement to composite ghost-free HDR images. To this end, dense correspondences have to been detected, for example, by PatchMatch method (HaCohen et al., 2011). These patch-based methods are time-consuming, because the computation of dense correspondences and the following joint optimization on image reconstruction is very expensive. Moreover, the quality of HDR images heavily relies on the choice of the reference image because the saturated pixels in the reference image yield big problem in patch matching (Hu et al., 2013; Zheng and Li, 2015; Zheng et al., 2013).

1.2. Issues in current HDR techniques and challenges in exposure-robust alignment

The standard and widely-used method for HDR imaging is to combine differently exposed LDR images of the same scene. The motivation behind this technique is that different exposures capture different dynamic range characteristics of the scene. However, real applications of this technology suffer from two problems (Srikantha and Sidibé, 2012): (1) Misalignment: global motion from hand-held camera results in misaligned images that cause the combined HDR image to look blurry; (2) Ghosting: ghost artifact appears due to dynamic scenes. To well cope with the two problems and generate high-quality HDR images, the choice of reference image is very important (Zheng et al., 2013). To our best knowledge, most of the HDR methods select the middle-exposed image, which is viewed as the best exposed image, as an initial reference, and subsequently processing (aligning and detecting moving objects) the consecutive images. After that, the processed images are served as new references to deal with their consecutive images until all images are done. The key advantage of the progressive method is to alleviate the difference of the underlying images resulting from different exposures. However, the drawback of the progressive solution is the error propagation. For example, two images with large EV interval have large registration error. This implies that the performances of the existing paradigms are limited by the number of images, which affects the quality of final HDR images. Hence, it is desirable to use a fixed reference and develop an exposure-robust method to align the multi-exposed images.

It is highlighted that, two images with large exposure value (EV) interval have significant intensity variations as shown in Fig. 1. Such images pose great challenges for image registration:

(1) A feature (e.g., an indoor feature) detected in one image (the long-exposed image) may not occur in another one (the short-exposed image) as shown in Fig. 1, which results in difficulty to use feature-based methods;
(2) Each image contains severely under/over-exposed regions, where information is loss and few features are detected;
(3) The intensities or intensity gradients are not linearly related with exposure times as supposed in Aguiar (2006); Bartoli (2008); Luong et al. (2010); Zimmer et al. (2011) due to nonlinear CRF property, which yields difficulty to normalize these images for area-based registration;
(4) A specific intensity in one image may map to multiple intensities in the other images, and vice versa. For example, two saturated pixels $Z_1(u) = 255,\ Z_1(v) = 255$ in a long-exposed image may become $Z_2(u) = 255,\ Z_2(v) = 240$ in its corresponding short-exposed image. Or one grayscale, say 150 in a short-exposed image, may map to grayscales of 190 or 193 because of noise or quantization error.

Due to the aforementioned issues, two images with large EV interval are usually not correlated. We used Photoshop CS5 to align two images with more than 7EV interval. Results show that Photoshop recognizes them as different scenes and displays “the images intended for alignment should overlap by approximate 40%”.

1.3. Contributions of this paper

To address the issue, i.e., the middle-exposed image is supposed to be the best quality image and selected as reference, then progressive method is employed for image registration, which results in error propagation, the aim of this paper is to provide a robust method so that the alignment performance does not rely on the particular reference image. In other words, even when there are large over/under exposed areas in the selected reference image, and the underlying images have large EV interval, the proposed scheme can still yield high-performance alignment results. The contributions of this paper are summarized as follows:

(1) The invariant representation of multi-exposed images is analyzed, and it is indicated that exposures change the intensities, but keep the relative order of intensities. However, saturation yield inconsistent binary features of multi-exposed images.

(a) Long-exposed image (b) Short-exposed image

Fig. 1. Differently exposed images.

(2) The order features of multi-exposed images are theoretically analyzed, and a novel mutual local ternary pattern (MLTP) is proposed to cope with large-variation intensities. It is analyzed that the proposed MLTP is a good representation to preserve consistency of differently exposed images.
(3) A fast and effective alignment is developed for image alignment. A linear model derived from Taylor expansion is presented to estimate motion parameters. To improve accuracy and efficiency, image rotation is initially detected by the histogram-based matching, and coarse-to-fine technique is implemented to cope with large movement.

In the following section, we explore the invariant representation of differently exposed images. A novel MLTP is proposed in Section 3. The alignment algorithm via linear optimization combined with coarse-to-fine technique and histogram-based matching is presented in Section 4. Section 5 demonstrates experimental results, followed by conclusions in the last section.

2. Invariant representation of differently exposed images

Image registration depends on two issues: invariant features of images, and invariant similarity measure. An invariant representation brings out common information among images, while suppressing non-common information. For a sequence of differently exposed images, which has large variations in intensity and contain large saturated pixels, the meanings are two-fold: (1) The sequence represents identical scene, therefore, the pattern (or feature) extracted from each image is expected to be consistent; (2) As the sequence is captured in different exposures, the pattern from each image cannot be exactly the same. Accordingly, invariant representation is a crucial problem in aligning these images.

The function of a digital camera is to convert scene radiance E into intensity image Z with a specific exposure $\triangle t$, which can be modeled as $Z = \Psi(E\triangle t)$, where Ψ is called CRF. The CRF is normally assumed to be a monotonically increasing function (Debevec and Malik, 1997). Suppose that a sequence of images $Z_k \in \mathbb{R}^{M\times N}(k = 1, 2, \cdots q)$ of the same scene with varying exposures are captured, and the CRF Ψ is known. An image with exposure j can then be converted into one with exposure k (Debevec and Malik, 1997). As such, the multi-exposed images can then be normalized into a specific exposure. Yet, CRF computation is not easy because it is a nonlinear function.

On the other hand, assuming that two images Z_1, Z_2 of the identical scene are captured with exposure times $\triangle t_1$, $\triangle t_2$ respectively, and $E(u)$, $E(v)$ are two radiances at positions u, v respectively, the following relationship is normally held due to the monotonic property of the CRF:

$$E(u) \geqslant E(v) \Rightarrow Z_1(u) \geqslant Z_1(v), Z_2(u) \geqslant Z_2(v) \tag{1}$$

where $Z_i(k)(i = 1, 2)$ are intensities at position k in ith image. This implies that exposures change the intensities, but keep the relative order of intensities. Accordingly, we can use the intensity order as feature to represent the differently exposed images.

Based on the above idea, Ward proposed a method called Median Threshold Bitmap (MTB) (Ward, 2003), which converts an image to a binary map by the median value. Due to its simplicity and fast speed, this scheme is widely used to cope with multi-exposed images (Grosch, 2006; Jacobs et al., 2008; Lu et al., 2009; Pece and Kautz, 2010). It is worth noticing that much information of the original images is lost by the simple median threshold. Moreover, the MTB method cannot guarantee exactly equal division because there are many pixels in each grayscale.

While the MTB method offers global conversion, other non-parametric ordering features, such as census transform (CT) (Zabih and Woodfill, 1994), local binary pattern (LBP) (Ojala et al., 1996), BRIEF (Calonder et al., 2012) have been proposed to provide pixel-wised local transformation, which preserves more information of the original images. However, saturation and quantization may yield inconsistent binary features of multi-exposed images instead of Eq. 1:

$$\begin{aligned} &E(u) > E(v) \Rightarrow Z_1(u) = Z_1(v), \ and \ Z_2(u) > Z2(v) \\ or \ \ &E(u) > E(v) \Rightarrow Z_1(u) > Z_1(v), \ and \ Z_2(u) = Z2(v) \end{aligned} \tag{2}$$

Eq. 2 indicates that the binary patterns using CT and LBP, are not robust to multi-exposed images.

A straightforward solution to alleviate the above problem is to use parametric ordering features, such as the local ternary patter (LTP) method (Tan and Triggs, 2010), a generalized LBP representation to tolerate small difference of intensities, or the correlation map (Akyüz, 2011), a generalized LTP feature which allocates different values in consideration of positions. Several LTP variants such as scale invariant local ternary pattern (SILTP) (Liao et al., 2010), extended local ternary pattern (ELTP) (Liao, 2010), and adaptive extended local ternary pattern (AELTP) (Mohamed and Yampolskiy, 2012), have been presented to improve robustness to noise and illumination change. In the following section, we reveal that the LTP and its variants are not robust in representing differently exposed images, and a novel MLTP is accordingly proposed.

3. Mutual local ternary pattern and its computation

3.1. Local ternary pattern

Consider a pixel at position u with intensity $Z(u)$, and $v(j)(j = 1, 2, 3, 4)$ are its 4-connected neighborhoods. The LTP string S of the pixel at position u is defined as follows (Tan and Triggs, 2010):

$$S^j(u,\delta) = \begin{cases} 1 & if \quad Z(v(j)) - Z(u) > \delta \\ 0 & if \quad |Z(v(j)) - Z(u)| \leqslant \delta \\ -1 & if \quad Z(v(j)) - Z(u) < -\delta \end{cases} \qquad j = 1,2,3,4 \quad (3)$$

where δ is a user-defined threshold. It is seen that the LTP, like the LBP, represents the relative intensity in its neighborhood rather than absolute intensity. By introducing the parameter δ, the non-parametric LBP string is replaced by the ternary string S. The results conducted on face recognition in Tan and Triggs (2010) show that LTP is more discriminant and less sensitive to noise in uniform regions than LBP.

While the δ is selected as a constant in Tan and Triggs (2010), some researchers argue that such choice does not account for illumination and noise level. In Liao et al. (2010), the SILTP descriptor was proposed to handle illumination variations, in which the intensities are assumed to follow scale transform. Instead of fixed threshold, Liao proposed extended LTP (ELTP) in Liao (2010), in which the value of δ is proportional to the standard deviation of the local patch. Such choice helps to retain invariance with respect to illumination transformation. Based on the work of Liao (2010), an approach called AELTP was introduced in Mohamed and Yampolskiy (2012) to adaptively compute each threshold by simple statistical operation. To improve the robustness to noise, a noise tolerant ternary pattern (NTTP) was proposed in Shoyaib et al. (2011). It is worth highlighting that the original LTP and its variants are determined by the underlying image without analyzing the relationship among image pairs. It is argued that these methods are valid only in a specific condition and a MLTP scheme is accordingly presented in the following section.

3.2. Mutual local ternary pattern

Let Z_1, Z_2 be two images with the identical scene in exposure time $\triangle t_1$, $\triangle t_2$ respectively. For two pixels at positions u, v respectively, the relative intensities are

$$\begin{aligned} \triangle_1(u,v) &= Z_1(v) - Z_1(u) \\ \triangle_2(u,v) &= Z_2(v) - Z_2(u) \end{aligned} \quad (4)$$

Assuming that the intensity relationship f between two images Z_1, Z_2 is available, i.e.: $Z_2 = f(Z_1)$, we have

$$\begin{aligned} Z_2(v) - Z_2(u) &= f(Z_1(v)) - f(Z_1(u)) \\ &= f(\triangle_1(u,v) + Z_1(u)) - f(Z_1(u)) \end{aligned} \quad (5)$$

The following equation is derived from Eq. 5 using Taylor expansion combining Eq. 4:

$$\triangle_2(u,v) = \triangle_1(u,v) f^{(1)}(Z_1(u)) + \sum_{n=2}^{\infty} \frac{\triangle_1^n(u,v)}{n!} f^{(n)}(Z_1(u)) \quad (6)$$

in which $f^{(n)}(Z_1(u))$ is the nth-order derivative in $Z_1(u)$. Eq. 6 describes the relationship of relative intensity between two images. We can draw the following conclusions based on Eq. 6:

1) $\triangle_1(u, v)$ is generally not equal to $\triangle_2(u, v)$. This implies that the images in different exposures have different contrasts (i.e., intensity difference $\triangle_i(u, v)$). Therefore, it is not reasonable to select a fixed threshold δ for a sequence of differently exposed images as Tan and Triggs (2010) does.
2) The intensity difference $\triangle_i(u, v)$ in image i is related with not only the intensity difference $\triangle_j(u, v)$ in image j, but also the derivatives $f^{(n)}(Z_j(u))(n = 1, 2, \ldots)$. Due to nonlinear property of function f, the intensity differences are not linear mapping.
3) If $\triangle_i(u, v)$ is small, Eq. 6 can be approximated by

$$\triangle_2(u,v) = \triangle_1(u,v) f'(Z_1(u)) \quad (7)$$

Remarks:

1) It is highlighted that the derivative $f'(Z)$ is discrete in digital images. The quantization by the imaging system and image processing (interpolation) may result in the fact: one gray-level in an image may map to several gray-levels in the other image and vice versa. In such situation, the derivative $f'(Z)$ is always zero and leads to $\triangle_2(u, v) = 0$, no matter how $\triangle_1(u, v)$ changes.
2) The LTP is particularly superior to the LBP in two cases:
   (1) Two under/over-exposed images, because saturation always results in situations in which the derivative $f'(Z)$ is zero;
   (2) The region is uniform, because the error from intensity quantization, interpolation or noise, may cause the positive values $\triangle_i(u, v)(i = 1, 2)$ to be negative and vice versa.

It has been analyzed that the relationship of intensity difference $\triangle_i(u, v)(i = 1, 2)$ depends on function f, which maps an intensity in one image to an intensity in another image. This function is called intensity mapping function (IMF). There are several methods to determine the IMF in situations where multi-exposed images of the same scene are captured, for example, the SVD method (Debevec and Malik, 1997), the comparagram method (Mann, 2000) and the cumulative histogram method (Grossberg and Nayar, 2003). While the former two methods require finding corresponding intensity pairs, the latter offers a flexible solution to determine the IMF, i.e., it finds the IMF even for images with both camera motion and dynamic scenes.

Suppose that Z_1 and Z_2 are two images with approximate radiances, $H_1(Z)$ and $H_2(Z)$ represent their corresponding cumulative histograms. It is shown in Grossberg and Nayar (2003) that the IMF $f_{12}(Z)$ mapping intensities in image Z_1 to image Z_2 can be estimated by the following equation

$$f_{12}(Z) = H_2^{-1}(H_1(Z)) \quad (8)$$

Similarly, we find the function $f_{21}(Z) = H_1^{-1}(H_2(Z))$, which maps intensities in image Z_2 to image Z_1.

While determining the parameter δ, we notice that the LTP derived from each image is expected to be consistent due to an identical scene. However, the LTP from each image cannot be exactly the same because two images are captured in different exposures. Such target implies that the value of δ is determined by two images instead of one image alone.

For image Z_1: after obtaining the IMF f_{12}, the intensity difference $\triangle_2(u, v)$ mapped in another exposure is predicted to be $\triangle_2(u,v) \approx \triangle_1(u,v) f'_{12}(Z_1(u))$. To avoid the uncertainty in which the intensity differences $\triangle_1(u, v)$ and $\triangle_2(u, v)$ are not related, the proposed MLTP of image Z_1 is determined as follows:

$$S_1^j(u) = \begin{cases} 2 & \begin{array}{l} if \quad Z_1(v(j)) - Z_1(u) > \delta_1, \\ and \quad f_{12}(Z_1(v(j))) - f_{12}(Z_1(u)) > \delta_2 \end{array} \\ 1 & \begin{array}{l} if \quad Z_1(v(j)) - Z_1(u) < -\delta_1, \\ and \quad f_{12}(Z_1(v(j))) - f_{12}(Z_1(u)) < -\delta_2 \end{array} \\ 0 & otherwise \end{cases} \qquad j = 1,2,3,4. \quad (9)$$

Similarly, the MLTP of image Z_2 is determined as follows:

$$S_2^j(u) = \begin{cases} 2 & \begin{array}{l} if \quad Z_2(v(j)) - Z_2(u) > \delta_1, \\ and \quad f_{21}(Z_2(v(j))) - f_{21}(Z_2(u)) > \delta_2 \end{array} \\ 1 & \begin{array}{l} if \quad Z_2(v(j)) - Z_2(u) < -\delta_1, \\ and \quad f_{21}(Z_2(v(j))) - f_{21}(Z_2(u)) < -\delta_2 \end{array} \\ 0 & otherwise \end{cases} \qquad j = 1,2,3,4. \quad (10)$$

In summary, the proposed MLTPs by Eqs. 9 and 10 have the following unique properties:

1) The MLTP of one image is not only determined by the image itself, but also by its counterpart;
2) The MLTP $S_i^j(u)$ is intensity-dependent, which is affected by the IMF. Due to the nonlinear property of IMF f_{12} and f_{21}, the value of $S_i^j(u)$ is adaptive to its mapping function at $Z_i(u)$ which is controlled by parameter δ_2.
3) Theoretically, selecting $\delta_2 > 0$ ensures that the MLTPs of two differently exposed images are consistent. If a larger value of δ_2 is selected, more $S_i^j(u)$ are equal to 0, which indicates that more pixels are categorized as uncertainty. The effects of parameters δ_1, δ_2 on pattern consistency and performance will be discussed in Section 5.

4. Alignment algorithm

4.1. Estimation of motion parameters

According to Eqs. 9 and 10, the multi-exposed images can be converted into MLTP-string images, in which each element contains ternary-string features. Then image alignment can be performed by conventional matching algorithms, for example, Hamming distance, used in the BRIEF method (Calonder et al., 2012). Our experiments show that the matching in three dimensions is time-consuming.

Alternatively, the MLTP-string images are uniquely represented by feature images $F_k \in \mathbb{R}^{M\times N}(k = 1, 2, \ldots, q)$, in which each element is coded as

$$F_k(u) = \sum_{j=0}^{3} S_k^j(u)3^j \tag{11}$$

Thus, the feature images, also known as MLTP-coded images $F_k(k = 1, 2, \ldots, q)$ can be regarded as normalized images of differently exposed images, in which the 3-value coded intensities are within the range of [0, 80], and accordingly the pixel-based methods can be employed for image alignment. In the following, a fast optimization method via a linear model is proposed.

Let F_1 and F_2 be the MLTP-coded images of two differently exposed images Z_1 and Z_2 respectively. Assuming that the rotational angle is θ, and the translations between Z_1 and Z_2 are T_x and T_y in x, y directions respectively, we have

$$F_2(x, y) = F_1(x\cos\theta - y\sin\theta - T_x, y\cos\theta + x\sin\theta - T_y) \tag{12}$$

If the rotation is small, we approximate $\sin\theta \approx \theta$, $\cos\theta \approx 1$. Expanding F_1 by Taylor series to the first order, the following linear equation yields (Keren et al., 1988):

$$\Phi \cdot P = K \tag{13}$$

where

$$\Phi = \begin{pmatrix} \sum\left(\frac{\partial F_1}{\partial x}\right)^2 & \sum\frac{\partial F_1}{\partial x}\frac{\partial F_1}{\partial y} & \sum R\frac{\partial F_1}{\partial x} \\ \sum\frac{\partial F_1}{\partial x}\frac{\partial F_1}{\partial y} & \sum\left(\frac{\partial F_1}{\partial y}\right)^2 & \sum R\frac{\partial F_1}{\partial y} \\ \sum R\frac{\partial F_1}{\partial x} & \sum R\frac{\partial F_1}{\partial y} & \sum R^2 \end{pmatrix} \tag{14}$$

$$K = \left(\sum(F_1 - F_2)\frac{\partial F_1}{\partial x}, \sum(F_1 - F_2)\frac{\partial F_1}{\partial y}, \sum R(F_1 - F_2)\right)^T \tag{15}$$

$$P = (T_x, T_y, \theta)^T \tag{16}$$

and $R = x\frac{\partial F_1}{\partial y} - y\frac{\partial F_1}{\partial x}$. The partial derivatives of an image are represented as a Gaussian smoothed image as follows:

$$\frac{\partial F_i}{\partial x} = -\frac{2x}{\delta^2}g(x, y)F_i(x, y),\ \frac{\partial F_i}{\partial y} = -\frac{2y}{\delta^2}g(x, y)F_i(x, y) \qquad i = 1, 2 \tag{17}$$

and $g(x, y)$ is the Gaussian function, i.e., $g(x, y) = \exp(-\frac{x^2+y^2}{\delta^2})$.

Therefore, the parameters $P = (T_x, T_y, \theta)^T$ are determined by linear regression:

$$P = \Phi^{-1} \cdot K \tag{18}$$

The solution can achieve sub-pixel accuracy. To improve accuracy, iterative procedure proposed in Keren et al. (1988) is implemented to solve Eq. 18.

It is noticed that the solution using the first-order approximation shown in Eqs. 13–16 is accurate only if motion between two images is very small. To register images with large movement, the coarse-to-fine technique (Bergen et al., 1992) and histogram-based matching are proposed.

4.2. Coarse-to-fine technique

The coarse-to-fine technique is represented by the Gaussian pyramid, which consists of low-pass filtered and down-sampled images of the preceding level of the pyramid. The base level is defined as the original image. Accordingly, for a MLTP-coded image $F(x, y) \in \mathbb{R}^{M\times N}$, the Gaussian pyramid in level l is defined as follows:

$$G_l(x, y) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} g(m, n)G_{l-1}(2x + m, 2y + n) \tag{19}$$

where $g(m, n)$ is the Gaussian kernel convolved with $G_{l-1}(x, y)$.

Assuming two images with translations T_x in x direction and T_y in y direction, and L-level Gaussian pyramids built, the translations T_x^L, T_y^L in the Lth level image $G_L(x, y)$ are

$$T_x^L = \frac{T_x}{2^L}, \quad T_y^L = \frac{T_y}{2^L} \tag{20}$$

Therefore, more levels lead to smaller translation. The level L is determined by the smallest image, which is not less than 32×32 in this work.

4.3. Rotation estimation by histogram-based matching

It is noted that the Gaussian pyramids alter image translations in different levels but do not affect image rotation. Large rotation results in poor performance of the linear model in Eq. 18. Accordingly, histogram-based matching is proposed to detect rotational angles before using a linear optimization.

Let $F_1, F_2 \in \mathbb{R}^{M\times N}$ be the feature images obtained from Eq. 15 respectively. For a feature image $F_i(i = 1, 2)$, the histogram $H_i^x \in \mathbb{R}^{80\times N}(i = 1, 2)$ in x direction is:

$$H_i^x(j) = (h_0^x(j), h_1^x(j), h_2^x(j), \ldots, h_{80}^x(j))^T, i = 1, 2, j = 1, 2, \ldots, N \tag{21}$$

in which $h_k^x(j)(k = 1, 2, \ldots, 80)$ is the count when element is k in a specific column j. The rotational angle θ^x is then computed as follows (Wu et al., 2014):

$$\Theta_x^*(x_0, \theta^x) = \underset{\theta,x}{argmin}\sum_{j=1}^{N}|H_1^x(j) - H_2^x(j, \Theta(x, \theta))| \tag{22}$$

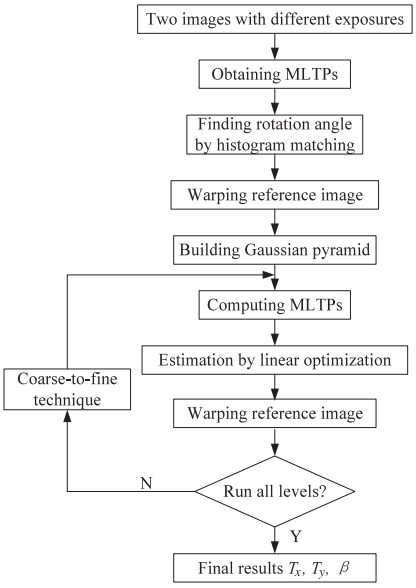


Fig. 2. The proposed scheme for aligning differently exposed images.

Similarly, we obtain the histogram $H_i^y \in \mathbb{R}^{80\times M}(i=1,2)$ by projecting the feature image $F_i(i=1,2)$ in y direction:

$$H_i^y(j) = \left(h_0^y(j), h_1^y(j), h_2^y(j), \ldots, h_{80}^y(j)\right)^T, i=1,2, j=1,2,\ldots,M \tag{23}$$

and the rotational angle θ^y is determined by:

$$\Theta_y^*(y_0,\theta^y) = \underset{\theta,y}{argmin}\sum_{j=1}^{M}|H_1^y(j) - H_2^y(j,\Theta(y,\theta))| \tag{24}$$

If $|\theta^x - \theta^y| < \zeta$, where ζ is a small value, it is guaranteed that the estimations in both x direction and y direction are correct, and the rotation angle is determined as follows:

$$\theta_0 = (\theta^x + \theta^y)/2 \tag{25}$$

It is seen that the histogram-based matching decouples the x, y spaces so that the searching time is $O(d^2)$ instead of $O(d^3)$. Set as an initial value for further optimization in Eqs. 13–18. The entire algorithm is described in Fig. 2.

5. Experimental results

In this section, the proposed method is evaluated by a variety of synthetic sequences which are benchmark in HDR imaging, and real sequences captured by NIKON D3/D300 or CANON EOS-1Ds Mark II which are available in Wu et al. Each sequence is arranged from long exposure to short exposure. The results of the proposed method are compared with (1) existing non-parametric ordering features using the MTB, LBP[1] methods; (2) LTP and its variants, such as SILTP (Liao et al., 2010), ELTP (Liao, 2010), AELTP (Mohamed and Yampolskiy, 2012) methods. (3) the SIFT method (Tomaszewska and Mantiuk, 2007), which is the best feature descriptors; (4) commercial products, such as Photomatix, the best software for HDR imaging, and Photoshop CS5, the most popular tool in image processing. It is noted that, while different ordering features are used to represent multi-exposed images, the same alignment algorithm shown in Section 6 is performed. All experiments were performed in Matlab environment. For the proposed method, the predefined parameters are selected as $\delta_1 = 2$ and $\delta_2 = 1$. For other parametric methods, such as SILTP, ELTP, AELTP etc., the recommended parameters by the authors are used. For the SIFT-related methods in Tomaszewska and Mantiuk (2007), the results are obtained via PSFTools.[2] As the commercial tools PSFTools, Photomatix[3] and Photoshop only provide aligned images without motion parameters, their performances are subjectively judged by overlaying the aligned images.

Table 1
Overall results performed on benchmark database.

Method	Mean of rotation error	Mean of translation error[a]
MTB	2.9	24.3
LBP	2.7	54.8
LTP	3.4	39.9
SILTP	3.8	43.9
ELTP	2.8	39.7
AELTP	2.7	25.0
Proposed MLTP	1.3	8.1

[a] The error is computed by both y-axis and x-axis translations

5.1. Tests on synthetic images

There are 9 sequences which are frequently-used for HDR imaging in the database as shown in Fig. 3. All the sequences are well geometrically registered. Most sequences are captured in 1EV interval but some, for example "Mask" and "WindowSeries", are taken in 2EV interval. The data sets are available in Wu et al., which is named as benchmark database.

To evaluate the alignment performance and robustness of the aforementioned invariant features, the first image in each sequence is selected as the reference, and other images are rotated by 5°, and shifted by 30 pixels and 10 pixels in y-axis and x-axis respectively. The alignment performances in terms of mean errors for the 9 sequences in different methods are tabulated in Table 1. The results show that the proposed method achieves the best performance in terms of motion errors.

As typical examples, the "Memorial" and "WindowSeries" sequences are good choices to test algorithm robustness to exposure because both cover very wide dynamic range. The "Memorial" sequence has 16 images taken in 1EV interval as shown in Fig. 4. The "WindowSeries" sequence has 9 images taken in 2EV interval (therefore, the total exposure is 18EV intervals) as shown in Fig. 5. The performances on the two sequences in comparison with other ordering features, feature-based method (SIFT), and the specialized HDR tools, are demonstrated in Tables 2–3. For easy observation, the shifts listed in Tables 2–3 are integers, although sub-pixel accuracy can be achieved.

It is seen from Table 2 that both the AELTP and the proposed MLTP achieve the best result for the "Memorial" sequence, i.e., robust to 10EV increments. Compared with the reference of "Memorial" sequence, the reference of the "WindowSeries" sequence is nearly saturated, in which the mean value is 244.5, and the median value is 255. The performances on this sequence by the LBP, LTP, SILTP are not good, while the MTB, ELTP, AELTP and the proposed method achieve 10EV increments. The overall performances shown in Table 1 and Tables 2–3 indicate that the proposed MLTP outperforms other ordinal features, and the performance is steadily robust to 10EV increments. It is also observed from Tables 2–3 that Photoshop is robust to 7EV increments, better than the PFSTools which implements the SIFT algorithm, whereas the performances of the Photomatix are very poor. Our experiments demonstrate that Photomatix is useful only for small movement.

[1] Matlab code: http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab.

[2] Available at http://pfstools.sourceforge.net/.

[3] Available at http://www.hdrsoft.com/.

Fig. 3. Data sets used in experiments (a: Belgium, b: BigTree, c: Mask, d: GrandCanal, e: CadikDesk, f: ChurchDesk, g: Memorial, h: SnowMan, i: WindowSeries).

Fig. 4. Memorial sequence.

Fig. 5. WindowSeries sequence.

Table 2
Performances on "Memorial" sequence.

Method\EV		2	3	4	5	6	7	8	9	10	11
MTB	$\triangle\theta$	0	0	0	0	0	0	0	23.5	3.6	17.6
	$\triangle T_y$, $\triangle T_x$	2, 1	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	47,92	10, 6	77, 51
LBP	$\triangle\theta$	0	0	0	0	0	0	0	0	5.2	0.8
	$\triangle T_y$, $\triangle T_x$	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	35, 7	6, 66
LTP	$\triangle\theta$	0	0.2	0	0.2	0	0	9.0	16.9	18.6	3.2
	$\triangle T_y$, $\triangle T_x$	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	39, 0	34,27	35, 33	24, 4
SILTP	$\triangle\theta$	0	0	0	0	0	0.6	2.0	0.6	3.5	3.3
	$\triangle T_y$, $\triangle T_x$	1, 1	1, 1	1, 1	1, 1	1, 1	5,76	14, 3	87,89	142,80	122,49
ELTP	$\triangle\theta$	0	0	0	0	0	0	0	0.1	7.5	4.9
	$\triangle T_y$, $\triangle T_x$	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	18, 86	51, 0
AELTP	$\triangle\theta$	0.3	0	0	0	0	0	0	0.9	0	0.2
	$\triangle T_y$, $\triangle T_x$	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	45, 2	1, 1	26, 5
Proposed	$\triangle\theta$	0.1	0.2	0	0.3	0	0	0.2	0	0.6	4.6
MLTP	$\triangle T_y$, $\triangle T_x$	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	1, 1	1, 0	3, 11
SIFT[c]		√[b]	√	√	√	√	×	×	×	×	×
Photomatix[c]		×	×	×	×	×	×	×	×	×	×
Photoshop[c]		√	√	√	√	√	√	N/A [a]	N/A	N/A	N/A

[a] Photoshop cannot do alignment as it detects that overlap of the two images is less than 40%.
[b] indicates good alignment, × indicates wrong alignmen
[c] Judged by overlaying the aligned images

Table 3
Performances on "WindowSeries" sequence.

Method\EV		2	4	6	8	10	12	14	16
MTB	$\triangle\theta$	1.1	0.5	1.1	0.4	0.2	10.1	8.3	6.3
	$\triangle T_y$, $\triangle T_x$	2, 1	0, 0	0, 1	0, 0	0, 1	100,96	9, 8	9, 216
LBP	$\triangle\theta$	3.6	0.6	0.2	2.0	1.6	7.4	5.8	6.2
	$\triangle T_y$, $\triangle T_x$	46, 5	131, 2	132, 9	78, 6	131, 6	56,186	37, 1	8, 200
LTP	$\triangle\theta$	0	0	2.3	7.0	7.0	2.1	3.6	3.2
	$\triangle T_y$, $\triangle T_x$	1, 1	1, 1	9, 1	21, 13	93,206	82, 63	75, 51	25, 48
SILTP	$\triangle\theta$	1.7	5.6	16.1	11.4	9.1	3.2	5.6	6.4
	$\triangle T_y$, $\triangle T_x$	3, 2	99, 92	106,73	11, 18	19, 11	6, 7	67, 76	72, 95
ELTP	$\triangle\theta$	0	0	0.1	0.1	0.1	3.4	3.9	6.1
	$\triangle T_y$, $\triangle T_x$	1, 1	1, 1	1, 3	1, 2	1, 3	16, 2	54,196	24, 202
AELTP	$\triangle\theta$	0	7.7	0	0	0	1.0	5.9	8.3
	$\triangle T_y$, $\triangle T_x$	1, 1	109,16	1, 2	1, 2	1, 2	1, 3	45, 5	44, 28
Proposed	$\triangle\theta$	0	0	0.1	0	0.3	6.0	10.8	6.98
MLTP	$\triangle T_y$, $\triangle T_x$	1, 1	1, 1	0, 2	0, 1	2, 4	27, 10	20, 9	0, 37
SIFT		√	√	√	√	√	×	×	×
Photomatix		×	×	×	×	×	×	×	×
Photoshop		√	√	√	√	√	×	×	×

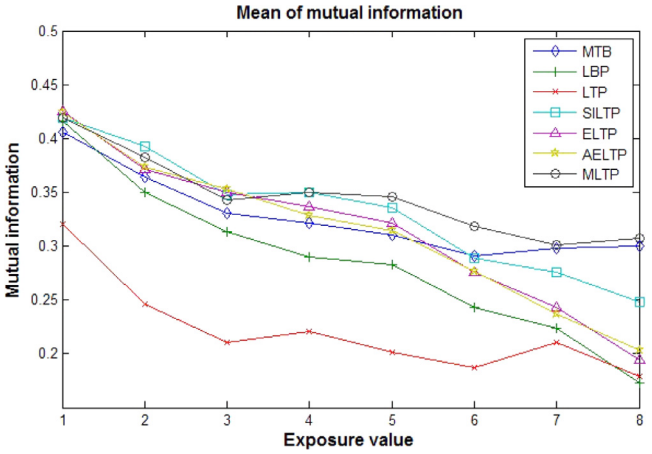


Fig. 6. Mutual information vs exposure value.

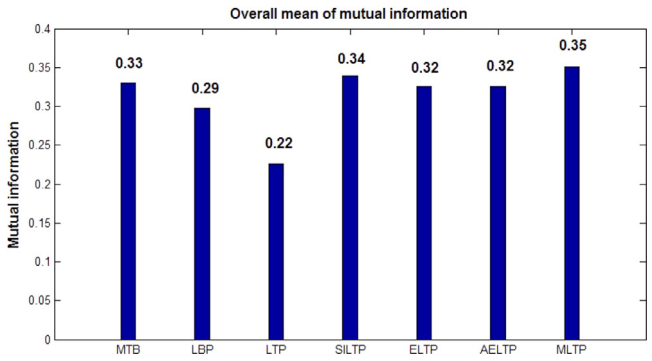


Fig. 7. Average mutual information based on the 35 real sequences.

Table 4
Operation times (second) in Matlab code.

Sequence	No. of images	Image size	MTB	LBP	MLTP
Great Canal	3	1025 × 661	18.6	13.7	27.0
Memorial	12	431 × 623	71.1	66.5	84.2
Snowman	6	693 × 505	72.8	59.4	72.7

5.2. Tests on real images

There are 35 real sequences captured using a NIKON D300, or a CANON EOS-1Ds Mark II, which comprise of various scenarios including small/big camera motion, static/dynamic scenes, little/severe saturation, 1EV/2EV intervals and so on. As motion parameters are unknown, mutual information (MI) (Viola and Wells III, 1997) is employed for alignment assessment, which is the best index for similarity measure of multi-exposed images (Wu et al., 2015). Selecting the brightest image as reference, the MIs in different exposures are shown in Fig. 6, and the average MIs based on the 35 real sequences are shown in Fig. 7. Fig. 6 reveals that (1) The performance normally decreases as EV interval increases; (2) The LTP does not offer better results than the LBP due to large intensity variation; (3) The MIs decrease greatly for the LBP, ELTP, AELTP and SILTP as EV interval increases, but not for the MTB and the proposed MLTP, which implies that the MTB and MLTP are robust to exposure. The overall performances based on the 35 real sequences shown in Fig. 7 indicate that the proposed MLTP achieves highest performance.

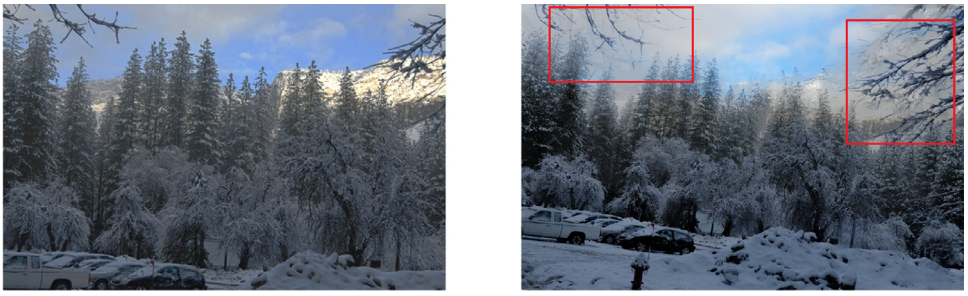
5.3. Tests on efficiency

To demonstrate efficiency of the proposed method, the operation times on Dell Precision T7400 are tabulated in Table 4, in comparison with the MTB and LBP methods. It is seen that the three descriptors have comparative efficiency.

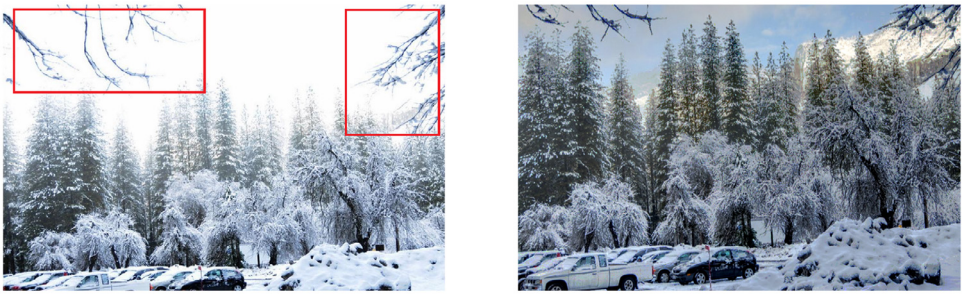
The algorithm of aligning differently exposed images mainly comprises of two parts: pattern extraction and pattern matching. Both the MTB and the LBP descriptors offer very efficient solutions for pattern extraction, whereas the MLTP method is expensive in computation because it needs to compute IMF before extracting the MLTP pattern. The operation time on pattern matching significantly depends on the features for approximation. Good patterns result in uniform convergence and fast approximation in estimating motion parameters. Generally speaking, the proposed MLTP method provides comparative running time although it is more complex, which, on the other hand, reveals that the patterns extracted by the MLTP method are more consistent and robust. It is noted that the experiments were performed in Matlab environment. The operation time can be further improved by optimizing the program and running in C code.

5.4. Results on HDR images

In this section, the proposed algorithm is employed for image alignment, and the exposure fusion in Li et al. (2014) combining with the ghost removal algorithms in Li et al. (2010); Zheng et al. (2013) are employed to provide ghost-free HDR images. The fusion scheme in Li et al. (2014) works in gradient



(a) Result by Photoshop (b) Result by Photomatix



(c) Result by patch-based method (d) Result by the proposed method

Fig. 8. HDR images for a static scene.

(a) Result by Photoshop (b) Result by Photomatix

(c) Result by patch-based method (d) Result by the proposed method

Fig. 9. HDR images for a dynamic scene.

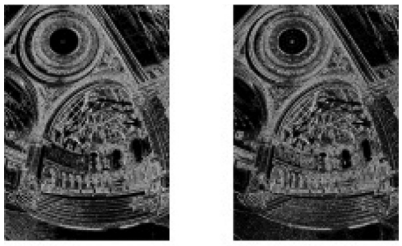
domain, which is common-used solution in multi-exposure composition (Ferradans et al., 2012; Zhang and Cham, 2012). Two results by the proposed method in real sequences (refer to Wu et al.), in comparison with those by the Photoshop, Photomatix, and the patch-based HDR reconstruction (Sen et al., 2012) are exemplified in Fig. 8 and Fig. 9.

In these experiments, the first (brightest) image is selected as reference for Photoshop, patch-based method and the proposed method. But the Photomatix can only choose the middle-exposed image as reference. For the static scene shown in Fig. 8, it is observed that blurring and ghost artifact (highlighted regions) are obvious for the Photomatix and the patch-based methods, which is yielded by misalignment of the Photomatix method and patch error by the patch-based method. The HDR images by the Photoshop and the proposed method are very clear which implies that the image sequence is well aligned. It is seen from Fig. 8(C) and Fig. 9(C) that the quality of HDR image is not good as there are large saturated regions in the final images. The reason is obvi-

Table 5
Mean errors $(\triangle\overline{\beta}, \triangle\overline{T})$ for the 9 sequences in different thresholds.

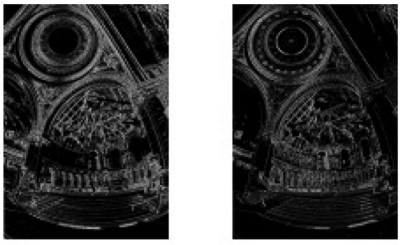
$\delta_2 \backslash \delta_1$	1	2	3	4	5	6	7
1	1.8, 6.3[a]	**1.3, 5.6**	1.7, 7.9	1.8, 13.0	2.0, 12.6	2.9, 18.7	2.5, 16.9
2	2.1, 10.1	1.5, 5.8	1.4, 6.8	1.7, 8.9	1.4, 7.8	2.5, 13.7	2.3, 15.5
3	2.5, 8.1	1.6, 6.2	1.5, 5.4	1.7, 6.3	1.6, 11.2	2.6, 14.5	2.2, 11.2
4	3.0, 11.6	1.7, 8.0	1.5, 8.1	1.8, 7.0	1.9, 12.8	2.6, 12.4	2.0, 12.1
5	3.8, 16.0	2.2, 9.5	1.7, 8.9	1.8, 8.2	1.9, 12.6	2.4, 10.9	2.1, 11.5

[a] The error $\triangle\overline{T}$ is computed by both y-axis and x-axis translations



(a) Image (1) (b) Image (5)

Fig. 10. MLTP-coded images in $\delta_1 = 2$, $\delta_2 = 1$.



(a) Image (1) (b) Image (5)

Fig. 11. MLTP-coded images in $\delta_1 = 8$, $\delta_2 = 2$.

ous: for the saturated regions in the reference, the corresponding regions in the transformed images by IMF method are still saturated. When we fused the image sequence by the code provided by Sen et al. (2012), it may yield the problem, as displayed "there is no invalid information from other images". The HDR images by the Photoshop and the proposed method are generally good as shown in Figs. 8 and 9. Compared to the results by the photoshop, the HDR images by the proposed method display more details, larger depth of field and richer colors.

5.5. Discussions

As mentioned in Section 1, multi-exposed images have large variations with many saturated pixels, which yield poor performance by using pixel-based method directly. The saturation also poses difficulty in using feature-based methods for feature matching. It is indicated in Section 2 that the ordinal measures, like MTB, LBP or the proposed MLTP, are reasonable approaches to represent invariance of differently exposed images. Here, we further provide insight on the ordinal representation of differently exposed images.

1) How do parameters δ_1 and δ_2 in MLTP affect performances?

The proposed MLTP uses parameters δ_1 and δ_2 to control the resulted MLTP patterns. The alignment errors for different values of δ_1 and δ_2 on the benchmark database are tabulated in Table 5. Figs. 10 and 11 show the MLTP-coded images for identical two images but in different thresholds. It is observed that more image features are preserved in smaller thresholds. Table 5 shows that the best thresholds are $\delta_1 = 2$, $\delta_2 = 1$.

2) How do MLTP-coded images change by different counterparts?

(a) Image (1) (b) Image (10)

Fig. 12. MLTP-coded images in large EV increments ($\delta_1 = 2$, $\delta_2 = 1$).

The proposed MLTP in Eqs. 9 and 10 reveals that the MLTP is affected not only by the thresholds, but also by the IMF determined by the underlying images. Using thresholds $\delta_1 = 2$, $\delta_2 = 1$, the MLTP-coded images of images (1) and (10) in the "Memorial" sequence are shown in Fig. 12. Compared with Fig. 10, which is the result of images (1) and (5) in identical thresholds, the MLTP patterns of the image (1) are different even the same thresholds are used. This implies that the proposed MLTP is adaptive to its counterpart. Figs. 10 and 12 show that two images with smaller EV interval preserve more image features.

3) How do features vary in different EV intervals?

The meanings of differently exposed images are two-fold: (1) The sequence represents identical scene, therefore, the feature (or pattern) derived from each image is expected to be consistent; (2) The sequence is captured in different exposures, the pattern from each image cannot be exactly the same. Then, how exposure affects the extracted features and how many consistent features can be achieved by the MLTP method?

Let $F_k \in \mathbb{R}^{M\times N}(k = 1, 2)$ be the feature images obtained using ordinal descriptors from two differently exposed images $Z_k \in \mathbb{R}^{M\times N}(k = 1, 2)$. The term C is defined as a consistent measure of feature images F_1, F_2 as follows:

$$C = \frac{\text{No. of pixels with same pattern in } F_1 \text{ and } F_2}{MN} \times 100\% \quad (26)$$

Then, the overall consistent rates of different descriptors in different EV intervals based on the benchmark database are shown in Fig. 13. It is observed that the features by the MTB method generally have high consistency and the features are robust to EV increments in comparison with other ordinal descriptors. It is noted that the feature consistency by the MLTP increases if the EV interval increases, which implies that the features derived by the proposed MLTP method is very good.

4) Where do inconsistent patterns happen?

Selecting the images (1) and (10) of "Memorial" sequence as example, the inconsistent features obtained by different descriptors are shown in Fig. 14, in which "white" represents inconsistent pattern. Our experiments show that most inconsistent patterns (76%) in the MTB method happen at the pixels where the intensities are around median value (i.e., median value $\pm$ 5) for images (1) and (10). One reason for such inconsistency arises from conversion by MTB. As the median value corresponds to many

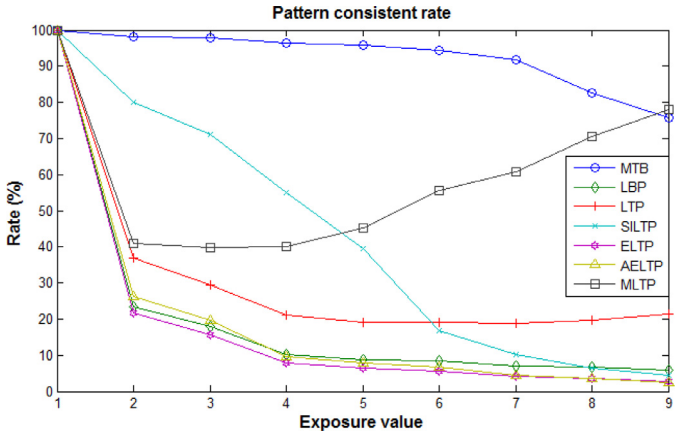
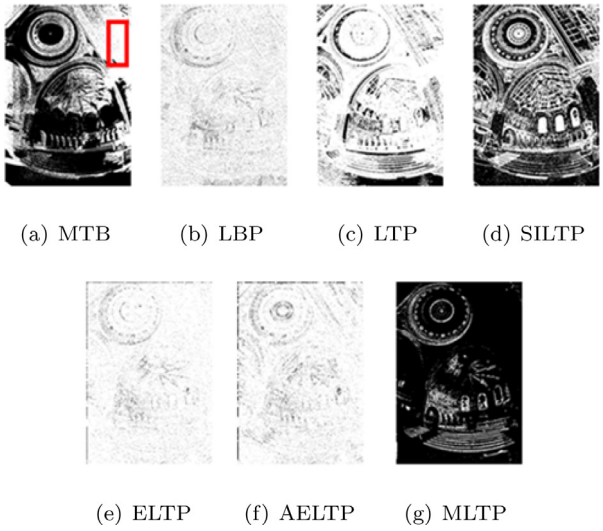


Fig. 13. Feature consistency in different EV intervals based on the benchmark database.



(a) MTB (b) LBP (c) LTP (d) SILTP

(e) ELTP (f) AELTP (g) MLTP

Fig. 14. Inconsistent patterns (in bright) by different descriptors.

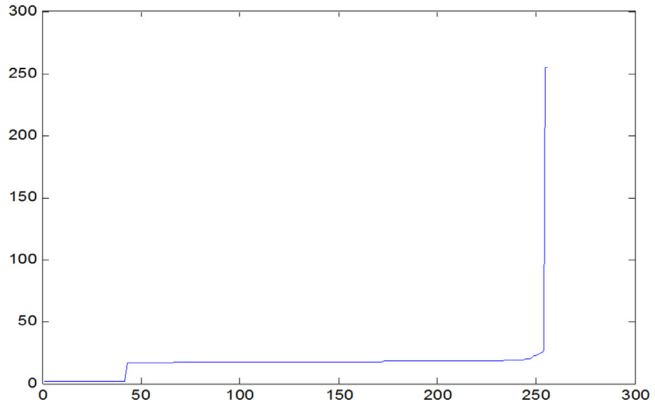


Fig. 15. IMF of images (1) and (10) in "Memorial" sequence.

pixels, the binary division cannot guarantee to be equal. The highlighted region in Fig. 14(a) shows such case in which the intensities are equal to median value. Another reason is due to nonlinear property of intensity mapping. Fig. 15 shows the IMF of images (1) and (10) in "Memorial" sequence. The nonlinear property of IMF leads to varying changes for each grayscale and results in inconsistent MTB patterns. Finally, noise is an important factor which yields pattern inconsistency. It is shown in Fig. 14 that pattern inconsistency happens everywhere for the LBP, LTP, ELTP and AELTP methods. Our experiment shows that 91.8% of the inconsistent patterns in the LBP method occur in the case that the relative intensities are within 5, which indicates that the LBP feature is not robust in smooth regions. Saturation is another source resulting in inconsistent features as shown in the SILTP method. It is observed from Fig. 14 that the inconsistent patterns in the proposed MLTP method mainly come from edges due to the following facts:

(1) Saturation which yields multi-value mapping as aforementioned;

(2) Error from the IMFs f_{12}, f_{21}. The functions f_{12}, f_{21} can be exactly obtained when the cumulative histograms $H_1(Z)$, $H_2(Z)$ are continuous functions. If $H_1(Z)$, $H_2(Z)$ are discrete functions, the IMFs are calculated according to:

$$\begin{aligned} f_{12}(Z) &= \underset{Z'}{argmin}\,|H_2(Z') - H_1(Z)| \\ f_{21}(Z) &= \underset{Z'}{argmin}\,|H_1(Z') - H_2(Z)| \end{aligned} \tag{27}$$

This implies that optimal but not accurate IMFs are obtained. Moreover, if the underlying images have large misalignment, the errors of IMFs by histogram matching are big. Hence, the IMFs are recomputed in each iteration as shown in Fig. 2.

6. Conclusions

Image registration depends on two issues: invariant features of images, and invariant similarity measure. As multi-exposed images used in HDR imaging have significant variations in intensity and contain large saturated pixels, aligning these images is challenging. In this paper, a novel mutual LTP method is presented to cope with differently exposed images. The proposed MLTP ensures consistent micro-structures of differently exposed images, which is photometrically invariant to monotonic transform of intensity function. Compared with other invariant representations, such as MTB, LBP, LTP etc., the proposed MLTP is determined by two images, which are independent of image saturation and robust to error arising from imaging system (for example, quantization) and image processing (quantization, interpolation etc.). A fast optimal scheme is developed to compute motion parameters. To solve large camera movement and large intensity variations accurately and robustly, the rotational angle is initially detected by histogram-based matching and the coarse-to-fine technique by Gaussian Pyramid is implemented. Our experimental results on 9 benchmarked multi-exposed sequences and 35 real images demonstrate that the proposed MLTP method is accurate and robust to 10EV increments, which outperforms other methods and current commercial HDR tools. The proposed method is very useful in situations whereby an image is selected and fixed as reference without progressive processing.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61371190. The authors wish to acknowledge the anonymous reviewers insightful and inspirational comments that have greatly helped to improve the technical contents and readability of this paper.

References

Aguiar, P.M., 2006. Unsupervised simultaneous registration and exposure correction. In: Proceedings of IEEE International Conference on Image Processing. IEEE, pp. 361–364.

Akyüz, A.O., 2011. Photographically guided alignment for HDR images. Eurograph. Areas Papers 73–74.

Bartoli, A., 2008. Groupwise geometric and photometric direct image registration. IEEE Trans. Pattern Anal. Mach. Intell. 30 (12), 2098–2108.

Bergen, J.R., Anandan, P., Hanna, K.J., Hingorani, R., 1992. Hierarchical model-based motion estimation. In: Proceedings of European Conference on Computer Vision. Springer, pp. 237–252.
Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P., 2012. BRIEF: Computing a local binary descriptor very fast. IEEE Trans. Pattern Anal. Mach. Intell. 34 (7), 1281–1298.
Candocia, F.M., 2003. On the featureless registration of differently exposed images. In: Proceedings of International Conference on Imaging Science, Systems & Technology. Citeseer, pp. 163–169.
Candocia, F.M., 2005. Analysis and enhancements to piecewise linear comparametric image registration. IEEE Trans. Image Proc. 14 (2), 181–188.
Cerman, L., Hlavac, V., 2006. Exposure time estimation for high dynamic range imaging with hand held camera. In: Proceedings of Computer Vision Winter Workshop. Czech Republic.
Debevec, P.E., Malik, J., 1997. Recovering high dynamic range radiance maps from photographs. In: Proceedings of ACM SIGGRAPH. ACM, pp. 369–378.
Eden, A., Uyttendaele, M., Szeliski, R., 2006. Seamless image stitching of scenes with large motions and exposure differences. In: Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition. New York, pp. 2498–2505.
Ferradans, S., Bertalmío, M., Provenzi, E., Caselles, V., 2012. Generation of HDR images in non-static conditions based on gradient fusion. In: Proceedings of International Conference on Computer Vision Theory and Applications, pp. 31–37.
Gevrekci, M., Gunturk, B.K., 2007. On geometric and photometric registration of images. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 1. IEEE, pp. 1261–1264.
Gevrekci, M., Gunturk, B.K., 2009. Illumination robust interest point detection. Comput. Vision Image Understand. 113 (4), 565–571.
Grosch, T., 2006. Fast and robust high dynamic range image generation with camera and object movement. In: Proceedings of Vision, Modeling and Visualization Conference, pp. 277–284.
Grossberg, M.D., Nayar, S.K., 2003. Determining the camera response from images: what is knowable? IEEE Trans. Pattern Anal. Mach. Intell. 25 (11), 1455–1467.
HaCohen, Y., Shechtman, E., Goldman, D.B., Lischinski, D., 2011. Non-rigid dense correspondence with applications for image enhancement. ACM Trans. Graph. 30 (4), 70.
Hossain, I., Gunturk, B.K., 2011. High dynamic range imaging of non-static scenes. In: Proceedings of IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics. 78760P–78760P.
Hu, J., Gallo, O., Pulli, K., 2012. Exposure stacks of live scenes with hand-held cameras. In: Proceedings of European Conference on Computer Vision. Springer, pp. 499–512.
Hu, J., Gallo, O., Pulli, K., Sun, X., 2013. HDR deghosting: How to deal with saturation? In: Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition. IEEE, pp. 1163–1170.
Jacobs, K., Loscos, C., Ward, G., 2008. Automatic high-dynamic range image generation for dynamic scenes. IEEE Comput. Graph. Appl. (2) 84–93.
Juergen, G., Rainer, G., 2009. Photographic Multishot techniques. Rocky Nook Inc, Santa Barbara, CA.
Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R., 2003. High dynamic range video. ACM Transactions on Graphics (TOG) 22 (3), 319–325.
Keren, D., Peleg, S., Brada, R., 1988. Image sequence enhancement using sub-pixel displacements. In: Proceedings of International Conference on Computer Vision & Pattern Recognition. IEEE, pp. 742–746.
Li, Z., Rahardja, S., Zhu, Z., Xie, S., Wu, S., 2010. Movement detection for the synthesis of high dynamic range images. In: Proceedings of IEEE International Conference on Image Processing. IEEE, pp. 3133–3136.
Li, Z., Zheng, J., Zhu, Z., Wu, S., 2014. Selectively detail-enhanced fusion of differently exposed images with moving objects. IEEE Trans. Image Process. 23 (10), 4372–4382.
Liao, S., Zhao, G., Kellokumpu, V., Pietikäinen, M., Li, S.Z., 2010. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1301–1306.
Liao, W.-H., 2010. Region description using extended local ternary patterns. In: Proc. of International Conference on Pattern Recognition. IEEE, pp. 1003–1006.
Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60 (2), 91–110.
Lu, P.-Y., Huang, T.-H., Wu, M.-S., Cheng, Y.-T., Chuang, Y.-Y., 2009. High dynamic range image reconstruction from hand-held cameras. In: Proceedings of IEEE Conference on Computer Vision & Pattern Recognition. IEEE, pp. 509–516.
Luong, H.Q., Goossens, B., Pižurica, A., Philips, W., 2010. Consistent joint photometric and geometric image registration. In: Proceedings of IEEE International Conference on Image Processing. IEEE, pp. 1197–1200.
Maintz, J.B.A., Elsen, P.A.V.D., Viergever, M.A., 1996. Comparison of edge-based and ridge-based registration of ct and mr brain images. Medical Image Anal. 1 (2), 151–161.
Mann, S., 2000. Comparametric equations with practical applications in quantigraphic image processing. IEEE Trans. Image Process. 9 (8), 1389–1406.
Mohamed, A.A., Yampolskiy, R.V., 2012. Adaptive extended local ternary pattern (aeltp) for recognizing avatar faces. In: Proceedings of International Conference on Machine Learning and Applications, 1. IEEE, pp. 57–62.
Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition 29 (1), 51–59.
Oldridge, S., Miller, G., Fels, S., 2011. Mapping the problem space of image registration. In: Proceedings of Computer and Robot Vision, pp. 309–315.
Pece, F., Kautz, J., 2010. Bitmap movement detection: HDR for dynamic scenes. In: Proceedings of Conference on Visual Media Production. IEEE, pp. 1–8.
Ramirez, R., Loscos, C., Martín, I., Artusi, A., 2013. Patch-based registration for auto-stereoscopic HDR content creation. In: Proceedings of 1st International Conference and SME Workshop on HDR Imaging.
Sen, P., Kalantari, N.K., Yaesoubi, M., Darabi, S., Goldman, D.B., Shechtman, E., 2012. Robust patch-based HDR reconstruction of dynamic scenes.. ACM Trans. Graph. 31 (6), 203.
Shoyaib, M., Abdullah-Al-Wadud, M., Chae, O., 2011. A noise-aware coding scheme for texture classification. Sensors 11 (8), 8028–8044.
Srikantha, A., Sidibé, D., 2012. Ghost detection and removal for high dynamic range images: Recent advances. Signal Process. Image Commun. 27 (6), 650–662.
Szeliski, R., 2006. Image alignment and stitching: a tutorial. Foundation. Trends Comput. Graph. Vision 2 (1), 1–104.
Tan, X., Triggs, B., 2010. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans. Image Process. 19 (6), 1635–1650.
Tico, M., Pulli, K., 2010. Robust image registration for multi-frame mobile applications. In: Proceedings of Asilomar Conference on Signals, Systems and Computers. IEEE, pp. 860–864.
Tomaszewska, A., Mantiuk, R., 2007. Image registration for multiexposure high dynamic range image acquisition. In: Proceedings of International Conference on Computer Graphics, Visualization and Computer Vision. Plzen, CzechRepublic.
Viola, P., Wells III, W.M., 1997. Alignment by maximization of mutual information. Int. J. Comput. Vision 24 (2), 137–154.
Ward, G., 2003. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. J. Graph. Tools 8 (2), 17–30.
Wu, S., Li, Z., Zheng, J., Zhu, Z., 2014. Exposure-robust alignment of differently exposed images. IEEE Signal Process. Lett. 21 (7), 885–889.
Wu, S., Li, Z., Zheng, J., Zhu, Z., 2015. Aligning multi-exposed images: What are the good feature and similarity measure? In: Proceedings of IEEE Conference on Industrial Electronics and Applications. IEEE, pp. 1959–1963.
Wu, S., Yang, L., Xu, W., Zheng, J., Li, Z., Fang, Z.,. A mutual local-ternary-pattern based method for aligning differently exposed images: supplementary materials. Available: https://sites.google.com/site/mltp2016wu/.
Zabih, R., Woodfill, J., 1994. A non-parametric approach to visual correspondence. In: Proceedings of European Conference on Computer Vision. Citeseer, pp. 151–158.
Zhang, W., Cham, W.K., 2012. Gradient-directed multiexposure composition. IEEE Transactions on Image Processing 21 (4), 2318–2323.
Zheng, J., Li, Z., 2015. Superpixel based patch match for differently exposed images with moving objects and camera movements. In: Proceedings of IEEE International Conference on Image Processing. IEEE, pp. 4516–4520.
Zheng, J., Li, Z., Zhu, Z., Wu, S., Rahardja, S., 2013. Hybrid patching for a sequence of differently exposed images with moving objects. IEEE Trans. Image Process. 22 (12), 5190–5201.
Zimmer, H., Bruhn, A., Weickert, J., 2011. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. In: Computer Graphics Forum, 30. Wiley Online Library, pp. 405–414.