夜视融合图像质量客观评价方法

张 勇[1,2],金伟其[1]

(1. 北京理工大学 光电学院光电成像技术与系统教育部重点实验室,北京 100081;
2. 军械技术研究所,河北 石家庄 050000)

摘 要:为建立通用、客观的融合图像质量评价方法,在分析图像质量评价与融合图像质量评价关系基础上,给出了图像质量评价与融合图像质量评价的一般表达式。依据信息理论和结构相似度评价方法,对建立的 4 种客观评价指标,采用 4 种融合方法获得的 36 幅融合图像进行了主观评价实验,统计分析结果显示,结合人类视觉系统的客观评价方法优于熵、交互信息量等评价指标,但仍未达到高度的主客观一致性,说明构建通用、高效、主客观一致性好的融合图像质量评价指标存在较大难度,同时对可能存在的原因进行了分析。

关键词:图像融合; 客观; 评价; 结构相似度

中图分类号:TP391.4 **文献标志码**:A **文章编号**:1007-2276(2013)05-1360-06

Objective assessment method of night vision fusion image quality

Zhang Yong[1,2], Jin Weiqi[1]

(1. Key Laboratory of Photoelectronic Imaging Technology and System, Ministry of Education of China, School of Optoelectronics, Beijing Institute of Technology, Beijing 100081, China; 2. Mechanical Engineering Institute, Shijiazhuang 050000, China)

Abstract: The aim of this paper is to study the practicability of the general and objective fusion image quality assessment method. Based on the analysis of the relationship between image quality assessment and fusion image quality assessment, the general expression of fusion image quality assessment was provided. Four fusion image evaluating metrics were presented based on information theory and structural similarity. Experiments were performed for 36 fusion images generated by 4 fusion algorithms fusing visible and infrared images. Experimental result shows that the objective evaluating method bonding to the human vision system is better than the metrics such as entropy and mutual information. However, the experiment result also show the assessment result is unsatisfied to the concordance subjective and objective assessing result, which explain that it is very difficult to build the fusion image evaluating metrics of the general, efficient and good consistency between subjective and objective result. Furthermore, the possible reason for performance assessment is discussed in the end.

Key words: image fusion; objective; assessment; structural similarity

收稿日期:2012-03-10; **修订日期**:2012-04-20

基金项目:总装备部基金

作者简介:张勇(1974-),男,高级工程师,博士生,主要从事融合图像质量评价方面的研究。Email:bit10701159@163.com

导师简介:金伟其(1961-),男,教授,博士生导师,博士,主要从事夜视技术方面的研究。Email:jinwq@bit.edu.cn

0 引 言

图像融合技术针对单一图像来源光学系统、成像机理、天候环境等因素限制,将不同类型传感器获取的信息进行有效综合和集成,一定程度上提高了图像承载信息量和空间分辨率。开展融合图像质量评价方法研究,不仅可比较融合算法的性能优劣,还可根据融合效果对系统进行改进和完善,具有重要的指导意义和参考价值[1-3]。

在分析图像质量评价与融合图像质量评价关系的基础上,提出了图像质量评价与融合图像质量评价的一般表达式,结合信息理论和结构相似度评价方法,建立了融合图像质量客观评价指标,并结合主观评价试验对其进行了验证。不失一般性,实验中选用的图像为可见光(微光)与长波红外灰度融合图像。

1 图像质量评价与融合图像质量评价

图像质量评价过程可理解为源图像经某种失真通道后降质程度的度量,融合图像相比于源图像,常可改善视觉效果,提高图像质量,可解释为源图像经某种增强通道后像质改善程度的度量;图像质量评价的源图像为单一图像,可作为标准参考图像与失真图像进行比较,融合图像质量评价的源图像为多幅图像,一般无标准参考图像(多聚焦融合系统除外);图像质量评价无特殊目的性,而融合图像质量评价常需结合实际应用的图像融合系统,带有明显的视觉任务或视觉目的。图像质量评价与融合图像质量评价可表示为:

$$\mathbf{Q}=\{\mathbf{T}(x,y),\mathbf{R}(x,y)\}_{\mathbf{E}} \tag{1}$$

式中:$\mathbf{T}(x,y)$,$\mathbf{R}(x,y)$分别为失真图像和输入图像,且有 $\mathbf{T}(x,y)=\mathbf{C}\cdot[\mathbf{R}(x,y)]+\mathbf{n}(x,y)$;$\mathbf{C}$ 为失真通道;$\mathbf{n}$ 为加性白噪声;$\mathbf{Q}$ 表示以 $\mathbf{E}$ 为图像质量评价指标获得的评价结果。针对融合图像质量评价,$\mathbf{T}(x,y)$表示融合图像,$\mathbf{R}(x,y)$为构造的实际并不存在的参考图像,融合图像与参考图像之间也由失真通道表示,代表融合图像在不同融合策略下向参考图像逼近。这里指的参考图像不同于图像质量评价中的标准参考图像,它是从多幅源图像中通过一定规则提取的信息或特征的集合,且能够对每幅源图像进行充分表示,而评价指标 $\mathbf{E}$ 则是尽可能的对这些信息或特征提取规则的具体表达。

2 融合图像质量客观评价方法

融合图像质量客观评价方法可分为无参考(Non-Reference,NR)或盲(Blind)评价方法、全参考评价方法(Full-Reference,FR) 和弱参考或减少参考(Reduced-Reference,RR)评价方法[4]。NR 评价方法摆脱了参考图像依赖,但需要抓住反映图像质量的最本质特征,而现有自然图像统计学模型还过于简单,构建一套通用目的的 NR 评价指标存在较大困难。FR 评价方法需借助标准参考图像实施,而实际应用的图像融合系统常无法获得,且在能够获得标准参考图像情况下,融合图像质量评价问题实际上退化为图像质量评价问题。RR 评价方法通过计算融合图像与源图像的信息量变化或典型特征差异来进行评判。文中依据信息理论和结构相似度评价方法,重点研究了 RR 评价方法。

2.1 基于信息理论的评价方法

熵评价(Entropy)是最基本的融合图像质量评价手段之一。在熵评价基础上,还衍生出了交叉熵、联合熵、偏差熵等一系列融合图像质量评价方法,其中交互信息量(Mutual Information,MI)应用最为广泛,可表示为:

$$MI_{FA}=\sum_{f=0}^{L-1}\sum_{a=0}^{L-1}p_{FA}(f,a)\log_2\frac{p_{FA}(f,a)}{p_F(f)p_A(a)} \tag{2}$$

$$MI_{FB}=\sum_{f=0}^{L-1}\sum_{a=0}^{L-1}p_{FB}(f,b)\log_2\frac{p_{FB}(f,b)}{p_F(f)p_B(a)} \tag{3}$$

$$MI_F^{AB}=MI_{A,F}+MI_{B,F} \tag{4}$$

式中:$p_A(a)$, $p_B(a)$, $p_F(f)$分别为源图像和融合图像的概率密度;图像 $\mathbf{A}$ 和 $\mathbf{B}$ 的联合概率密度为 $P_{AB}(a,b)$。

2.2 空域结构相似度评价方法

Wang Z 提出的结构相似度评价方法(Structural Similarity,SSIM)最初用于压缩图像质量评价[5],计算时采用滑动窗口方法,设 $\mathbf{X}$,$\mathbf{Y}$ 为待比较两幅图像的窗口图像块,对所有图像块评价结果求均值后得到两幅图像的平均结构相似度评价指标,可表示为:

$$MSSIM(x,y)=\left(\frac{1}{NM}\right)\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}SSIM_{x,y}(i,j) \tag{5}$$

SSIM 方法通过比较标准参考图像和失真图像的结构差异获得评价结果,属于 FR 评价方法。若将 SSIM 方法应用于融合图像质量评价,根据公式(1),

在假设 SSIM 方法有效前提下，用于融合图像质量评价的 FMSSIM 方法可表示为：

$$FMSSIM=\left(\frac{1}{NM}\right)\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}SSIM(f_i,\{I_{x,i}(m,n),I_{y,i}(m,n)\}_M) \quad (6)$$

式中：f 为融合图像；$I_{x,i}(m,n)$和 $I_{y,i}(m,n)$分别为输入源图像在同一位置图像块；M 为参考图像构造方法。常用构造方法以比较输入源图像块像素统计指标作为参考图像块的选择依据，如熵、均方根、对比度、尖锐度[6-7]等，构造过程中一般利用权重协调输入源图像在参考图像中的比例关系。

2.3 小波域结构相似度评价方法

空域内 SSIM 评价方法对平移、缩放、旋转或其他类型的未对准等几何失真具有高度敏感性，而这些失真往往由图像采集设备造成，并不是由于视觉场景中的物体结构变化所引起的[8]。为将 SSIM 评价方法应用于实际的图像融合系统，提出小波域结构相似度评价方法，具体过程如下：在小波域内对输入源图像和融合图像分别进行小波分解，同一个小波子带下同一空间位置提取的小波系数分别为 $c_x=\{c_{x,i}|i=1,2,\cdots,N\}$，$c_y=\{c_{y,i}|i=1,2,\cdots,N\}$，$c_f=\{c_{f,i}|i=1,2,\cdots,N\}$。根据公式(1)，融合图像质量评价需构造实际并不存在的参考图像，参考图像越能反映输入源图像的有效信息或典型特征，获得的评价结果越接近人眼的主观感受。图像低频部分表征图像滤除细节后的基本信息，将其加权平均作为参考图像低频子带的小波系数，高频部分表征图像的大部分细节特征，绝对值大的小波系数对应着显著的亮度变化，选择绝对值大的小波系数作为参考图像高频子带的小波系数。参考图像低频部分 $c_{r1,i}$ 和高频部分 $c_{r2,i}$ 的小波子带构造方法为：

$$c_{r1,i}=(c_{x,i}+c_{y,i})/2 \quad (7)$$

$$c_{r2,i}=\{|c_{x,i}|,|c_{y,i}|\}_{max} \quad (8)$$

利用和空域 SSIM 评价方法相似的表达式表示小波域结构相似度评价方法，用于融合图像质量评价的小波域结构相似度评价指标为：

$$\tilde{S}(c_r,c_f)=\frac{2\left|\sum_{i=1}^{N}c_{r,i}c_{f,i}^{*}\right|+K}{\left|\sum_{i=1}^{N}c_{r,i}\right|^2+\left|\sum_{i=1}^{N}c_{f,i}\right|^2+K} \quad (9)$$

式中：K 为小的非负常量；c* 表示 c 的复共轭。

为更好的实现评价结果的主客观一致性，引入人眼视觉敏感度带通特性，利用对比度敏感函数(Contrast Sensitivity Function，CSF) 将人眼最为敏感的中频区域赋予较大权重，将低频和高频部分赋予较小权重，设定不同频段的权值为 λ_i，则基于人类视觉系统(Human Vision System，HVS)的融合图像小波域结构相似度评价指标(DWTFMSSIM)为：

$$DWTFMSSIM(c_r,c_f)=\sum_{i=1}^{N}\lambda_i\cdot\frac{2\left|\sum_{i=1}^{N}c_{r,i}c_{f,i}^{*}\right|+K}{\left|\sum_{i=1}^{N}c_{r,i}\right|^2+\left|\sum_{i=1}^{N}c_{f,i}\right|^2+K} \quad (10)$$

3 实验分析

3.1 实验设计

实验目的是考核提出的客观评价指标的适用性以及与主观评价结果的一致性。实验图像共 9 幅，源于 MIT、TNO、美国海军实验室等单位在网络上(Http://www.imagefusion.org)公布的可见光(微光)与红外图像。融合算法包括取大平均(MAX)、主成分分析(Principal Component Analysis，PCA)、拉普拉斯塔型分解(Laplacian Pyramid，LP)和离散小波变换(DWT)。参与实验的观察者为 3 名男性、3 名女性，年龄均在 23~30 岁之间，有正常或经过校正后恢复正常的视力。实验前对测试者进行了初步训练，包括测试目的和评分依据。要求在无视觉任务的情况下，根据对融合图像的主观感受，给出图像整体的“感知质量”[9]。实验利用一台显示器，当源图像的图像尺寸过大而不能和融合图像同时显示时，观察者可通过图像切换对源图像和融合图像进行比较，在视觉无法确定两幅图像质量优劣时，尽可能采用相同的分数来避免分数的不确定性。评分依据见表 1。

表 1 评价尺度与分值

Tab.1 Assessment metric and score value

Subjective score	Image quality
7	
6	Good
5	
4	Fair
3	
2	Poor
1	

3.2 客观评价指标

利用 Entropy、MI、FMSSIM、DWTFMSSIM 作为客观评价指标,其中 FMSSIM 利用图像块标准差作为选择参考图像块的判据,即如果输入图像 **A** 中的图像块 **a** 的标准差大于输入图像 **B** 中对应位置图像块 **b** 的标准差,则参考图像块由 **a** 决定,反之则由 **b** 决定。DWTFMSSIM 利用 sym8 小波对输入源图像和融合图像进行 5 层分解得到 6 个频带,用 **LL** 表示最低频带,**H**1、**H**2、**H**3、**H**4、**H**5 表示各分解级的高频带,取各频带内 CSF 曲线的平均值,即 λ_i=[1.79,2.35,2.87,3.16,2.56,1.00][10]。

3.3 客观评价结果分析

图 1 中,图(a)~(d)分别为 MAX、PCA、LP、DWT 融合算法获得的融合图像。图 2 为 4 种评价指标针对不同融合图像的评价结果,为便于比较,对数值进行了归一化处理。计算 4 种评价指标的标准差,Entropy

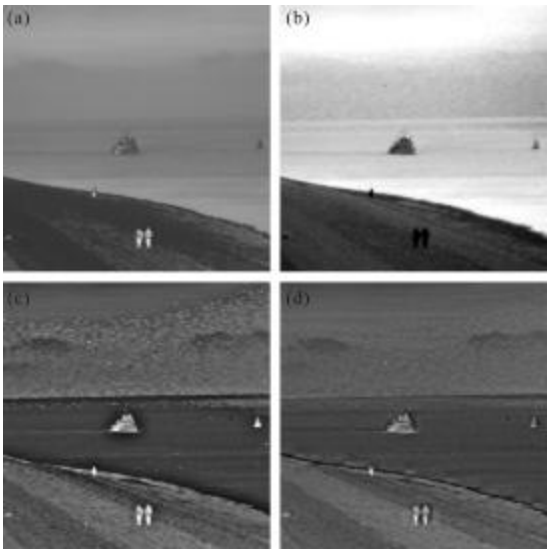

图 1 不同融合算法获得的融合图像

Fig.1 Fusion images by different fusion algorithms

为 0.06 1 1,MI 为 0.305 6,FMSSIM 为 0.208 5,DWTFMSSIM 为 0.129 5,可见 MI 分类效果最优,熵评价分类效果最差。结合图 2,熵评价除了对第 3、6、9 幅图像 PCA 算法评分较高外,其他评价结果呈现均衡性,说明熵评价指标离散程度偏弱,无法有效区分融合图像质量。MI 评价指标对 36 幅融合图像质量的评价结果倾向性明显,即对 MAX 和 PCA 算法给予较高分值,LP 和 DWT 算法分值较低。FMSSIM 评价指标分类程度适中,同时给予了 LP 和 DWT 算法较高分值,这与其算法注重图像结构特征变化存在关联。和 FMSSIM 指标相比,DWTFMSSIM 指标虽然提高了 LP 和 DWT 算法得分,但分类性能偏弱。

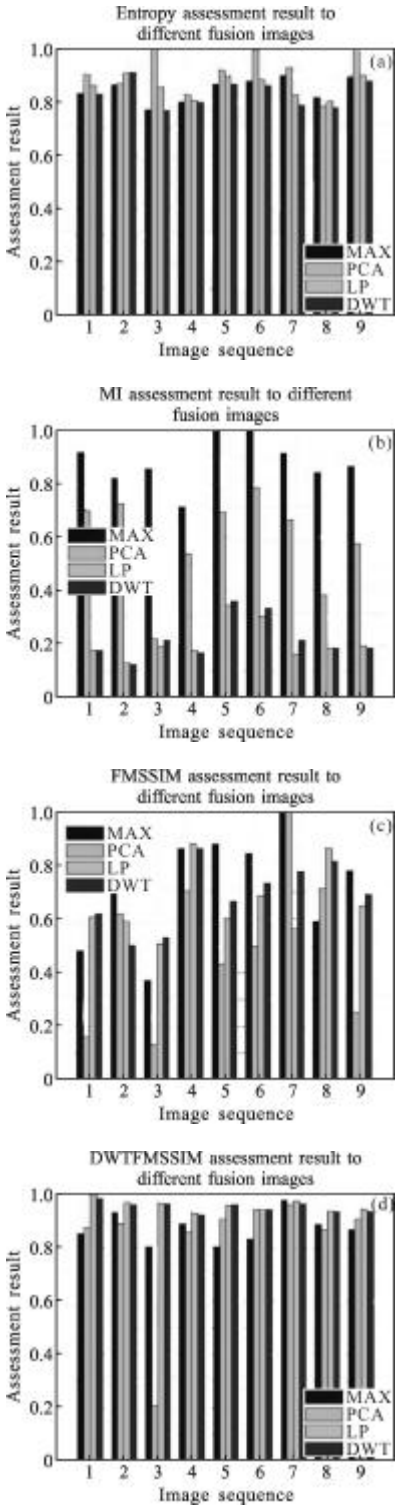

图 2 不同指标评价结果

Fig.2 Result using different evaluation metrics

3.4 主客观评价结果分析

为验证 4 种客观评价指标的主客观一致性，图 3 给出了图像整体“感知质量”与 4 种客观评价指标的相对关系。为将所有数据作相关性统计分析，对原始数据做 Z 变换(Z-score)。

$$Z_{score}=\frac{x_i-\bar{x}}{\sigma} \tag{11}$$

式中：x_i 为变量值；$\bar{x}$ 为变量均值；σ 为样本方差。

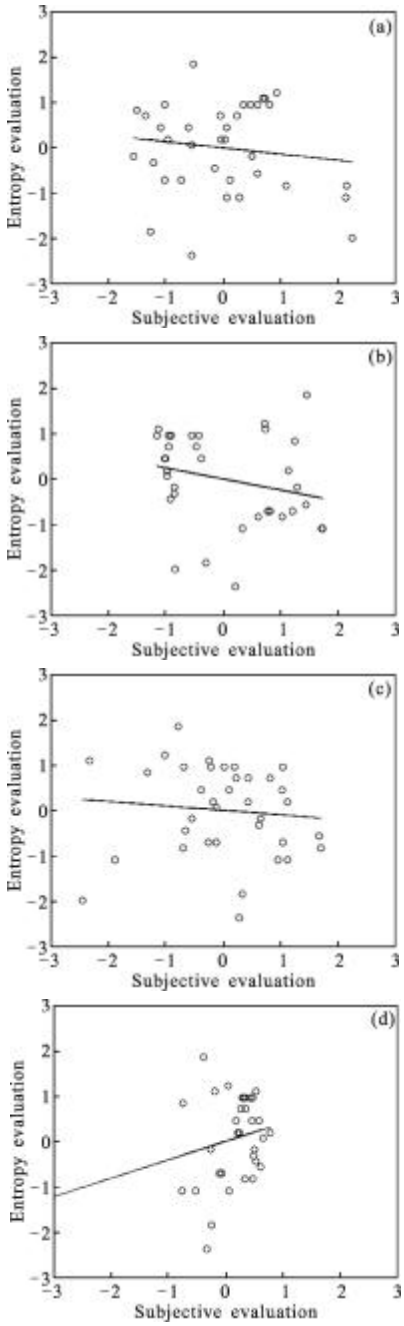

图 3 感知质量与客观评价指标关系

Fig.3 Relationship between perceptual quality and objective evaluating metrics

图 3 中，直线是线性回归结果，“感知质量”与 Entropy、MI、FMSSIM、DWTFMSSIM 等 4 种客观评价指标的相关系数分别是-0.1381，-0.2451，-0.0989 和 0.4080。可见结合 HVS 仿生模型的 DWTFMSSIM 评价方法与主观评价一致性较好，但从相关系数可知，4 种评价指标均未获得高度的主客观一致性。尤其值得关注的是，SSIM 方法在图像质量评价中得到广泛应用，但在融合图像质量评价方面却表现一般。分析其原因有以下四个方面：一是主观评价实验的样本集数量、人眼评判的随机性等因素可能导致主观评价结果的随意性，主观评价得分未能反映融合图像的“真正”质量；二是自然图像内容的复杂性，参与实验的图像包括舰船、车辆、人员等目标信息，水岸、林地、城市夜景等场景信息，且不同场景下图像亮度、对比度、纹理各不相同，人眼很难依据评分规则给出真正客观的分值；三是融合算法的局限性。不同融合算法针对不同的图像内容，融合效果存在较大差异，要求观察者针对不同的图像内容给出图像整体“感知质量”分值具有较大困难；四是融合算法的目的性。图像质量评价中的失真通道作用于图像全局，一般随影响程度的加强，图像失真程度逐步加大，如 JPEG、JPEG2000、Gaussian blur、Fast fading、白噪声等，根据公式(1)，融合图像在不同融合策略下向参考图像的逼近由不同的融合算法实现，不同的融合算法从不同角度对融合图像产生解释，产生不同亮度、不同对比度的融合图像，在提高人眼对图像某一部分视觉刺激的同时，可能也会降低人眼对图像其他部分的敏感度，如 MAX 算法在突出红外目标信息的同时，也会降低对图像大部分场景的感知能力，PCA 算法增强了灰度梯度大的图像成分，但也会受到盲元(Dead pixel)、噪声等灰度突变点的影响，使效果变差。LP 算法在不同分辨率上提取每幅源图像的显著特征，并有效结合这些特征得到融合图像，DWT 算法在分解过程中引入了空间方向，在抑制噪声的同时可有效提取图像的显著纹理和边缘特征，但是两种方法均可能降低人眼对目标信息的感知能力，正是因为融合算法设计阶段的目的性和侧重性，决定了融合图像质量评价较图像质量评价具有更大难度。

4 结 论

根据以上实验结果得出以下结论：

(1) 融合图像质量评价属于图像质量评价研究范畴,但相比于图像质量评价,构建通用、高效、主客观一致性好的融合图像质量评价指标存在相当大难度, 原因包括对人眼生理特性和心理特性建模的局限性, 也包括实际应用融合系统视觉任务或视觉目的的多样性。

(2) 图像整体“感知质量”作为人眼感知图像质量的最终主观评价分值过于笼统, 执行过程中常会受到观察者主观感受限制,对于不同的视觉任务,人眼主观评价过程必然会不自觉的引入既定的视觉目的,从而影响最终的评价结果。

(3) 考虑到考核客观评价指标性能的主要判据是主客观一致性, 下步工作应解决主观评价图像样本、观察人员和条件的不确定性对实验结果的影响,建立可用于融合图像质量主观评价的图像数据库,规范主观评价过程、方法、约束条件和数据处理过程,在构建通用评价指标未获突破的情况下,应结合实际应用的图像融合系统, 将明显的主观目的引入到客观评价指标中, 构建面向具体任务的融合图像质量客观评价指标。

参考文献:

[1] Zhang Yong, Jin Weiqi. Subjective assessment method of night vision fusion image quality [J]. Infrared and Laser Engineering, 2013, 42(2): 528-532. (in Chinese)
张勇, 金伟其. 夜视融合图像质量主观评价方法 [J]. 红外与激光工程, 2013, 42(2): 528-532.

[2] Yu Xianchuan, Pei Wenjing. Performance evaluation of image fusion quality metrics for the quality of different fusion methods [J]. Infrared and Laser Engineering, 2012, 41(12): 3416-3422. (in Chinese)
余先川, 裴文静. 针对不同融合算法的质量评价指标性能评估[J]. 红外与激光工程, 2012, 41(12): 416-422.

[3] Wang Haihui, Peng Jiaxiong, Wu Wei. Comparison of the methods of evaluating the performance of multi-sensor image fusion [J]. Infrared and Laser Engineering, 2004, 33(2): 189-193. (in Chinese)
王海晖, 彭嘉雄, 吴巍. 评价多传感器图像融合效果方法的比较[J]. 红外与激光工程, 2004, 33(2): 189-193.

[4] Zhou Wang, H R Sheikh, A C Bovik. No-reference perceptual quality assessment of JPEG compressed images [C]//IEEE International Conference on Images Processing, 2002, 1: I477-I480.

[5] Zhou Wang, A C Bovik, H R Sheikh, et al. Image quality assessment: from error visibility to structural similarity[C]// IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[6] Zheng Yongan, Song Jianshe, Zhou Wenming, et al. Objective evaluation method for image fusion based on image quality index [J]. Systems Engineering and Electronics, 2006, 28(3): 463-466. (in Chinese)
郑永安, 宋建社, 周文明, 等. 基于图像质量因子的图像融合客观评价方法[J]. 系统工程与电子技术, 2006, 28(3): 463-466.

[7] Piella G. A general framswork for multiresolution image fusion: from pixels to regions[J]. Information Fusion, 2003, 4(4): 259-280.

[8] Zhou Wang, Simoncelli Eero P. Translation insensitive image similarity in complex wavelet domain[C]//IEEE International Conference on Image Processing, 2005, II: 573-576.

[9] Shi Junsheng, Jin Weiqi, Wang Lingxue. Study on perceptual evaluation of fused image quality for color night vision[J]. J Infrared Millim Waves, 2005, 24(3): 236-240. (in Chinese)
石俊生, 金伟其, 王岭雪. 视觉评价夜视彩色融合图像质量的实验研究[J]. 红外与毫米波学报, 2005, 24(3): 236-240.

[10] Ding Xuxing, Zhu Rihong, Li Jianxin. A criterion of image quality assessment based on property of HVS[J]. Journal of Image and Graphics, 2004, 9(2): 190-194. (in Chinese)
丁绪星, 朱日宏, 李建欣. 一种基于人眼视觉特性的图像质量评价[J]. 中国图像图形学报, 2004, 9(2): 190-194.