# 夜视融合图像质量主观评价方法

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,在分析图像质量主观评价方法基础上,研究了融合图像质量主观评价对象、评价条件、评价指标和数据处理等关键环节,结合目标探测识别与场景理解两个典型的"视觉任务",提出了"目标可探测性"和"细节可分辨性"两个主观评价指标,并研究了图像整体"感知质量"与这两个指标的相关性。对 3 种不同场景的微光与红外融合图像,采用 9 种融合方法获得的 189 幅融合图像的统计分析结果显示,"细节可分辨性"和图像整体"感知质量"相关性好,而"目标可探测性"虽然和图像整体"感知质量"相关性较差,但在基于具体视觉任务的融合图像质量评价过程中,仍可作为有效评价指标之一。

**关键词:** 评价; 主观; 图像融合; 感知质量

**中图分类号:** TP391.4 **文献标志码:** A **文章编号:** 1007-2276(2013)02-0528-05

# Subjective assessment method of night vision fusion image quality

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

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

**Abstract:** In order to build the general and objective fusion image quality assessment method, the key technique of fusion image quality subjective assessment were presented such as assessment object, assessment condition, assessment index and data processing based on the analysis to image quality subjective assessment method. Two subjective assessment indexes were provided including target detectable ability and details resolvable ability according to two typical vision tasks of target detection and scene understanding. Furthermore, the correlation was investigated between the whole perceptual quality with target detectable ability and details resolvable ability respectively. Experiments were performed for 189 fusion images generated by 9 kinds of fusion algorithms fusing visible and infrared images under 3 kind different scenes. Statistical result showed that the high correlation between the whole perceptual quality with details resolvable ability compared with the low correlation to target detectable ability. However, the index of target detectable ability was still effective and could be used in the actually vision task.

**Key words:** assessment; subjective; image fusion; perceptual quality

---

**收稿日期:** 2012-06-20; **修订日期:** 2012-07-28

**基金项目:** 总装备部基金资助课题

**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工作。
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工作。Email:jinwq@bit.edu.cn

## 0 引 言

图像融合将不同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[1-2]。目前研究多集中在融合图像质量客观评价方法上[3]。融合图像质量主观评价是直接利用观察者对被测图像的主观反应来确定系统性能的一种测试方法,由于费时费力,长期以来未受重视,主观评价过程的随意性和评价指标的多样性限制了融合算法的发展,也给客观评价指标的建立带来困难。为此,借鉴图像质量主观评价方法和标准,从评价对象、评价条件、评价指标和数据处理方面对主观评价过程进行了研究, 然后利用实际图像库, 结合目标探测识别与场景理解两个典型的"视觉任务"或"视觉目的",提出了"目标可探测性"和"细节可分辨性"两个主观评价指标,并研究了图像整体"感知质量"与这两个指标的相关性[4]。不失一般性,实验中做如下约束:(一) 不考虑实际应用中图像处理、配准等因素对融合图像质量的影响;(二) 研究的图像为可见光(微光)与长波红外灰度融合图像;(三) 无标准参考图像。

## 1 主观评价实验设计

借鉴图像质量主观评价方法, 将融合图像质量主观评价实验分为四个环节:(一) 评价对象;(二) 评价条件;(三) 评价指标;(四) 数据处理。

### 1.1 评价对象

评价对象指参与融合图像质量主观评价的图像数据库。针对图像质量评价的数据库较为完备,一些公开的数据集可供使用,VQEG 的 LIVE 还给出了图像的差异平均意见得分 (Difference Mean Opinion Score, DMOS),非常方便研究者使用。而针对融合图像质量评价的数据库还需进一步完善, 目前只有 MIT、TNO、美国海军实验室、空军实验室等单位将少数实验图像公开给研究者使用。在构造融合图像评价数据库时应注意以下三个方面:(一) 不同传感器摄取的可见光(微光)与红外图像应在时间和空间上严格配准;(二) 相同条件下摄取的图像应保持一定数量;(三) 图像内容应尽量丰富以便验证融合图像相对于源图像的像质改善程度。

### 1.2 评价条件

评价条件指参与评价过程的观察者、设备及环境。

#### 1.2.1 观察者

参与实验的观察者分为受过训练和未受过训练两种。一般来说,受过专门训练的观察者能够比未受过训练的观察者更能注意到图像质量的细微变化[5],但是受过训练的观察者往往更关注于图像质量变化的特定缺陷, 从而产生和未受过训练观察者在图像质量判断方面的差异[6],大量实验中采用的观察者未经训练,只是在评价前对其进行初步训练,帮助观察者理解测试目的和评价分数的确定。为使主观评价在统计上有意义,参加评分的观察者人数不易过少,且最好能在年龄、职业、受教育程度上保持相对一致性,一般要求观察者性别有适当比例,具备正常或矫正的视力, 对于彩色夜视融合图像的主观评价还需要观察者具有正常的颜色视觉。

#### 1.2.2 设备

一般使用 CRT 或 LCD 显示器,也有采用头盔显示。实验前应使用查找表线性校正显示器,调整显示器亮度、对比度满足人眼观察舒适,对观察距离和视角没有统一要求。图像质量主观评价分为连续单刺激质量评价(Single Stimulus Continuous Quality Evaluation, SSCQE)、连续双刺激质量尺度(Double Stimulus Continuous Quality Scale, DSCQS)和比较双刺激尺度 (Double Stimulus Comparison Scale, DSCS),三种方式各有优缺点,如 DSCQS 方法对实验过程中图像损伤顺序对评价结果的影响较小, SSCQE 方法则对质量监控方面的评价更具代表性[7]。针对融合图像质量主观评价,主观评价的实验设计直接影响客观评价指标的建立和验证。SSCQE 方法摆脱了参考图像依赖,但评价过程中需抓住反映图像质量变化的最本质特征,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无参考(Non-reference, NR)或盲(Blind)评价方法较为困难。双刺激方法通过比较融合图像与源图像的差异或像质改变给出评价分值,一些客观评价指标的建立也是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,实验心理学传统结果表明,当测试感觉时,相对判断比绝对判断更加稳定、准确。在国际电联(ITU-R)BT.500 建议的众多评价方法中, 由于双刺激方法采用了标准参考图像,其结果比单刺激方法具有更高的灵敏度和稳定性。采

用 SSCQE 方法只需将融合图像显示于一台显示器上,而采用双刺激方法,由于受图像尺寸大小限制,输入源图像和融合图像未必能同时显示于一台显示器上,而缩放则会降低图像分辨率,影响评价效果。若采用多台显示器同时显示源图像和融合图像,则应尽量使用相同型号的显示器,且统一设置显示器分辨率和刷新频率。

#### 1.2.3 环境

测试一般在微暗房间内进行,测试过程中应保证观察者的舒适性,观察者一般采用双目观察,采用简单的计算机软件控制图像的显示、每幅图像的测试时间、图像转换等。对于目标探测任务,计算机软件还需记录观察者对目标的定位是否准确[8]。

### 1.3 评价指标

由于图像质量评价中的失真模型较为单一,且针对这些失真模型的建模较为成功,观察者在主观评价过程中利用对图像整体“感知质量”作为主观评价指标即可完成评价任务,且图像质量评价一般不涉及具体的视觉任务,而融合图像质量评价一般针对而实际应用的图像融合系统,常带有明显的视觉任务或视觉目的,如以同源光学传感器实现多聚焦场景理解以改善图像分辨能力、以异源光学传感器实现昼夜观测以增强目标探测识别能力等[9]。排除图像预处理和图像配准对融合图像的影响,不同的融合算法从不同角度对融合图像产生解释,从而产生不同亮度、不同对比度的融合图像。如加权融合算法直接在空域内对多幅源图像的对应像素点进行加权处理,主成分分析方法可增大灰度梯度大的图像成分,在处理两幅灰度梯度相差很大的图像时效果显著,塔型变换融合算法在不同分辨率上提取每幅源图像的显著性特征,并有效结合这些特征得到融合图像,和塔型变换融合算法相比,小波变换融合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。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,同时,不同融合算法针对不同的图像内容,融合效果也存在较大差异,这些都使得评价过程中要求观察者对不同的图像内容给出对图像整体“感知质量”的分值带来较大困难。考虑到局部统计特性具有空间不变性,且能够克服全局统计图像特征常表现出的明显的空间不稳定性,同时利用人眼在典型观察距离上,在给定瞬间仅对图像局部区域给予高分辨率感知的视觉机制特点,结合目标探测识别和场景理解两个具体的视觉任务,提出“目标可探测性(Target Detectable Ability,TDA)”和“细节可分辨性(Details Resolvable Ability,DRA)”两个主观评价指标[10]。“目标可探测性”定义为观察者观察图像时首次注意的具有一定认知含义的图像区域,如图像中的人员、车辆、舰船等。对于点源目标,观察者不仅应注重目标是否可探测,还应观察目标与背景的对比度关系,即目标的“突出性”,对于扩展源目标,观察者还应增加目标的细节与纹理分析,即细节程度越高,目标可探测性分数越高。“细节可分辨性”定义为观察者对明显目标区域之外的图像场景的理解能力,观察者应参照源图像观察融合图像保留源图像的场景、信息、细节、纹理、对比度等能力。针对两个主观评价指标的评分依据见表1。

表 1 评价尺度与分值

Tab.1 Assessment metric and score value

| Subjective Score | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|
| Image quality | - | Good | - | Fair | - | Poor | - |

### 1.4 数据处理

由于主观评价是一个漫长的过程,一般采用事后数据处理方法。为了减轻观察者情绪、动机、知识背景等因素对评价结果的影响,可在一定规则条件下对原始数据进行取舍。如通过设置置信区间来抛弃针对同一幅图像的主观得分奇异值;也可根据同一观察者主观得分奇异值次数确认其不符合评价资格,舍弃其全部评价分值[11]。

## 2 实验与分析

### 2.1 实验设计

实验图像一部分来源于美国麻省理工学院林肯实验室、荷兰 TNO 人力资源研究所、美国海军研究生院在网络上提供的可见光(微光)和红外图像,另一部分为利用实验室搭建的双传感器图像融合系统采集的可见光(微光)和红外图像如图 1 所示,图 1 中,图(a)和图(b)分别为利用该系统拍摄的部分可见光(微光)和红外图像(限于篇幅,其他图像不再列出)。

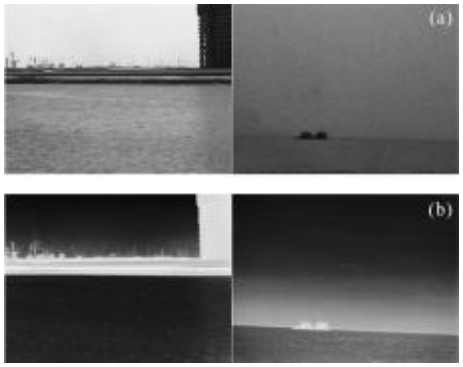


图 1 利用实验室搭建的图像融合系统获得的可见光与红外图像

Fig.1 Visible and infrared image gathered using actual image fusion system

参与实验的图像包括了 7 幅山林背景、7 幅海天背景和 7 幅城镇建筑物背景，不同背景条件下的典型目标包括人员、车辆、舰船等。针对每幅图像采用的融合算法包括取大平均法(MAX)、取小平均法(MIN)、均值法(AVG)、主成分分析法(PCA)、Laplacian pyramid (LAP)、Gradient pyramid (GRA)、DWT、CWT、SIDWT 等 9 种[12]，共获得融合图像 189 幅。

实验目的要求观察者给出不同融合图像的整体“感知质量”、“目标可探测性”和“细节可分辨性”三个主观评价指标。测试人员由实验室研究生共 14 人组成，其中 8 位男性，6 位女性，年龄在 23~30 岁之间，他们有正常或经过校正后正常的视力。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”。“感知质量”定义为视觉刺激引起的人的“舒适性主观感觉”，由于该指标反映为一个复杂的观察者心理过程，因此未对其进行严格约束，只是要求观察者对图像整体效果根据自身的心理感受给出合适的分数。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。实验中选用的显示器为 22 in (in=2.54cm)Samsung P2250W 型，分辨率设置为 1 024×768，刷新频率为 85 Hz，观察者和屏幕显示距离约为 0.6m，观察者也可根据自身条件进行适当调整。在视觉无法确定两幅图像质量优劣时，尽可能的采用相同的分数来避免分数的不确定性，同时要求测试者在不受判断时间限制的条件下尽可能准确的给出测量分值。

## 2.2 指标分析

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)。

$$z=\frac{x_i-\bar{x}}{\sigma} \tag{1}$$

式中：$x_i$ 为变量值；$\bar{x}$ 为变量均值；$\sigma$ 为样本方差。

图 2~图 4 分别给出了“感知质量”、“目标可探测性”、“细节可分辨性”的相对关系。图 2~图 4 中的直线是线性回归结果，数字是相关系数。“感知质量”与“目标可探测性”、“细节可分辨性”的相关系数分别是 0.820 7 和 0.647 4，可见“感知质量”与“细节可分辨性”的相关性较高，“细节可分辨性”可作为融合图像质量评价的主要指标。“感知质量”与“目标可探测性”的相关性较低，其原因除了实验图像均具有明显的目标可探测性外，目标所占图像区域远低于图像背景区域面积也是造成相关性较小的主要原因之一，因为即使是在目标足够引起主观感受的情况下，目标之外区域的“细节可分辨性”无明显的质量改观，“感知质量”也不能达到较高的得分。“目标可探测性”与“细节可分辨性”两个指标之间的相关系数为 0.621 9，由于不同融合算法在生成融合图像时，在

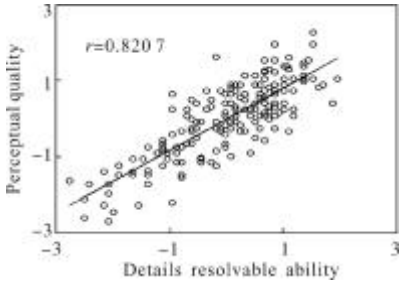


图 2 感知质量和细节可分辨性关系

Fig.2 Relationship between perceptual quality and details resolvable ability

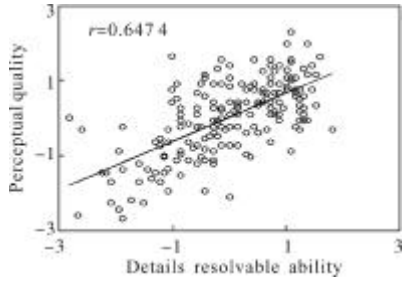


图 3 感知质量和目标可探测性关系

Fig.3 Relationship between perceptual quality and target detectable ability

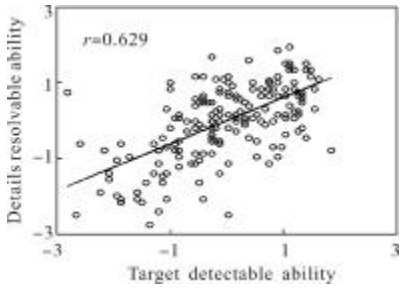


图 4 目标可探测性和细节可分辨性关系

Fig.4 Relationship between target detectable ability and details resolvable ability

改善目标可探测性的同时，往往也会提高图像对比度、纹理等细节特征，因此“目标可探测性”与“细节可分辨性”存在弱相关性，但两个指标仍是从不同角度评价融合图像质量。根据图 2~图 4，可构建基于“目标可探测性”和“细节可分辨性”的融合图像“感知质量”一般表达式，即：

$$PQ=k_1TD+k_2DR-\Delta \tag{2}$$

式中：**PQ** 为图像整体的感知质量；**TD** 为目标可探测性；**DR** 为细节可分辨性；Δ 为 **TD** 与 **DR** 之间的相关性；$k_1$，$k_2$ 为权重因子。通过调整两个指标的权重，可将公式(2)用于面向具体“视觉任务”或“视觉目的”的融合图像质量评价，当针对场景理解或目标探测识别任务时，公式(2)可改写为：

$$PQ_{visiontask}=\{TD\}\text{or}\{DR\} \tag{3}$$

## 3 结 论

根据文中实验结果得出以下结论：

(一) 提出的“目标可探测性”、“细节可分辨性”两个评价指标可用于融合图像质量主观评价。“细节可分辨性”与图像整体的“感知质量”相关性较高，“目标可探测性”与图像整体的“感知质量”相关性较低，但针对具体的“视觉任务”或“视觉目的”，“目标可探测性”仍可作为有效评价指标之一。

(二) 提出的“目标可探测性”、“细节可分辨性”两个评价指标可用于融合图像质量客观评价，从而方便验证主客观评价结果的一致性。利用目标自动识别算法结合感兴趣区域 (Region of interest，ROI)，可对源图像进行区域分割，从而实现“目标可探测性”指标的客观计算；利用结构相似度计算模型，充分考虑融合图像亮度、对比度及结构信息的变化情况，从而实现“细节可分辨性”指标的客观计算。

(三) 针对具体的“视觉任务”或“视觉目的”开展融合图像质量评价，可避免观察者对融合图像整体“感知质量”评价过程中带来的不稳定性，同时也符合实际应用的图像融合系统的设计目的，因为任何一种图像融合系统的产生均会和某种或几种视觉任务相关。

需要说明的是，实际应用的图像融合系统往往带有明显的视觉任务或视觉目的，这给融合图像质量评价带来了更大困难，但全面、客观的融合图像质量评价算法将是未来的发展方向，下步工作将考虑融合系统不同的应用范围和任务要求，扩展图像库，验证“目标可探测性”、“细节可分辨性”两个评价指标的客观计算方法。

**参考文献：**

[1] Zhang Junju, Chang Benkang, Zhang Baohui, et al. Long-distance image fusion system for infrared and LLL/visible bands [J]. Infrared and Laser Engineering, 2012, 41(1): 20-24. (in Chinese)

[2] Shi Shiming, Wang Lingxue, Wang Xia, et al. Real-time fusion system for color night vision [J]. Infrared and Laser Engineering, 2006, 35(s): 360-365. (in Chinese)

[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)

[4] toet A, frenken E M. Perceptual evaluation of different image fusion schemes [J]. Displays, 2003, 24(1): 25-37.

[5] Gote Nyman, Jenni Radun, Tuomas Leisti, et al. What do users really perceive-probing the subjective image quality [C]//SPIE, 2006, 6059: 605902.

[6] Cui C. Do experts and naïve observers judge printing quality differently [C]//SPIE, 2004, 5294: 132-145.

[7] Pinson M, Wolf S. Comparing subjective video quality testing methodologies [C]//SPIE, 2003, 5150: 573-582.

[8] Lanir J, Maltz M. Analyzing target detection performance with multispectral fused images [C]//SPIE, 2006, 6312: 631205.

[9] Lanir J, Maltz M, Rotman S R. Comparing multispectral image fusion methods for a target detection task [J]. Optical Engineering, 2007, 46(6): 0664021-0664028.

[10] 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)

[11] Zhou Jingchao, Dai Ruwei, Xiao Baihua. Overview of image quality assessment research [J]. Computer Science, 2008, 35(7): 1-5. (in Chinese)

[12] Vladimir Petrovc. Multi-level image fusion [C]//SPIE, 2003, 5099: 87-96.