

基于亮度和对比度模型的图像融合质量评估标准

李珊珊¹ 王琦² 冯兴乐¹

(¹ 长安大学电子信息工程系, 陕西 西安 710064; ² 长安大学交通信息与控制工程系, 陕西 西安 710064)

摘要 针对图像融合质量客观评估问题中, 结构相似度标准的视觉模型过于简化视觉特性的缺陷, 引入新的亮度和对比度模型, 提出一种新的图像融合质量评估标准。实验证明, 新的评估标准更能反映人的视觉感知与电脑显示硬件的特性, 与人的视觉感知标准更为接近。

关键词 图像处理; 图像融合质量评估; 结构相似度标准; 亮度; 对比度

中图分类号 O436 **文献标识码** A **doi:** 10.3788/LOP48.061001

A Novel Quality Metric Based on Luminance and Contrast Model for Image Fusion

Li Shanshan¹ Wang Qi² Feng Xingle¹

¹ Department of Electric Information Engineering, Chang'an University, Xi'an, Shaanxi 710064, China;
Department of Traffic Information and Control Engineering, Chang'an University,
Xi'an, Shaanxi 710064, China

Abstract In the objective quantitative evaluation without ideal image for different image fusion schemes, the vision model of structure similarity index measure metric oversimplifies the characteristic of vision perception and does not consist with monitor properties. We employ new luminance and contrast model and design a new image fusion performance evaluation metric. The experiments prove that the new metric takes more environment influence and monitor properties into account and better compliant with subjective evaluation.

Key words image processing; object performance evaluation for image fusion; structure similarity index measure; luminance; contrast

OCIS codes 100.0100;150.0150;070.0070

1 引言

图像融合是数据融合中一个非常重要的分支, 是一种综合了传感器、图像处理、信号处理、计算机和人工智能。Pohl 等^[1]将图像融合定义为对互补或冗余信息进行集合的过程。融合后的图像可信度更高, 模糊更少, 可理解性更好, 更符合人的视觉感知特性, 并能满足图像分割、特征提取、目标识别等图像处理功能的需要。图像融合通常在 3 个层次上进行: 像素级、特征级、决策级。最直观的评估融合效果的方法就是主观测试, 因为对于图像融合而言, 人是融合结果的最终评测者, 所有评估方法给出的评估结果都希望与人的视觉感受尽可能一致。Petroviè 等^[2]就用主观测试的方法评估融合结果的好坏。主观测试虽然简单、直接, 但其也有诸多不足之处, 例如耗时、昂贵、不能量化和优选某些融合系数, 此外, 人的视觉特性或心理状态都会影响评估结果, 这些都限制了主观测试法的应用。另外, 由于图像融合应用的场合与目的千差万别, 故而要求参与主观评价的观察者具备相当的专业知识水平。同时, 如果遇到待处理数据量庞大的情况, 则更会增加主观测试的难度。因此, 利用计算机自动进行图像融合质量的客观评估就成为了新的研究热点。

收稿日期: 2010-11-01; **收到修改稿日期:** 2011-02-18; **网络出版日期:** 2011-05-19

基金项目: 国家自然科学基金(60975028)、高等学校博士学科点专项科研基金(20090205120002); 和中国博士后科学基金(20080441159)资助课题。

作者简介: 李珊珊(1982—), 女, 博士, 讲师, 主要从事图像处理与分析、计算机视觉等方面的研究。

E-mail: liss@chd.edu.cn

如果我们能够得到一幅理想状态下的融合后图像,那么,只需要比较用不同算法获取的融合图像与这幅理想图像之间的相似程度,就可以知道它的质量好坏了。Liu 等^[3]列举了理想图像比较法常用的一些度量。主要有均值平方根误差(RMSE),标准最小均方误差(NLSE),信噪比峰值函数(PSNR),相关函数(CORR),差异熵(DE),以及交互信息(MI)。Zhang 等^[4]使用交互信息(MI)来度量融合图像的质量。MI 在信息论中被认为表达了从信源接收到的符号集中获取的关于信源发送的符号集的信息量(反之亦然)。在图像融合中使用这一概念,也就是可以计算融合后图像从源图像中获得的信息量。使用交互信息的问题在于,从数学角度分析,MI 量的最大值是当融合图像为所有源图像的均值时所取得的,但很显然,均值法绝对不是最好的融合方法,事实上在很多情况下它往往是最差的。这说明,MI 的度量结果与实际效果有不小的差异。Wang 等^[5]定义了结构相似度量标准(SSIM),也就是一种图像质量指数。Piella 等^[6]将它用于融合图像的质量评估。这种方法对于不在同一级进行的融合方法进行效果评估时效果不够理想。Li 等^[7]将输入图像的局部颜色作为融合后应保存的信息,计算融合图像与比较“重要”的源图像之间的局部颜色相似性,再得到全副图像的统计值,作为融合质量评价标准。

本文着眼于基于结构相似性的图像融合质量评估,并且指出这一类方法的缺点,引入更符合视觉感知的亮度和对比度模型,提出一种新的基于结构相似度的图像融合质量评估方法,并通过实验验证了这种新方法的有效性。

2 结构相似度标准的视觉模型的分析

SSIM 用于融合图像的质量评估时,具体的做法如下^[8]:

假设 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 和 $\mathbf{y} = (y_1, y_2, \dots, y_n)$ 是两个实数值序列, \bar{x} 是 \mathbf{x} 各元素的均值, δ_x^2 是 \mathbf{x} 各元素的方差,对 \mathbf{y} 亦然; δ_{xy} 是 \mathbf{x} 和 \mathbf{y} 的协方差,那么这两个向量的 SSIM 指数就是

$$Q_0 = \frac{\delta_{xy}}{\delta_x \delta_y} \cdot \frac{2\bar{x}\bar{y}}{\bar{x}^2 + \bar{y}^2} \cdot \frac{2\delta_x \delta_y}{\delta_x^2 + \delta_y^2}, \quad (1)$$

式中第一项是 x 和 y 的相关系数,取值范围为 -1 到 1 ;第二项是亮度项,衡量这两个代表图像的向量的亮度相似性,在图像中,所有的像素值均为正值,因此取值范围 $[0, 1]$;第三项是对比度项,衡量对比度分布的相似性,取值范围同样是 $[0, 1]$ 。且仅当两个向量完全相等时, Q_0 取到最大值 1 。

由于图像信号一般来说是不固定的,因此如果不是用来和理想图像比较,全局图像质量指数在实际应用中并不适合。因此 Wang 等^[5]又使用了固定大小的滑动窗口,从图像的最左上角逐像素移动直至最右下角,计算每个窗口中的本地质量指数,最后求均值

$$Q_0(a, b) = \frac{1}{|W|} \sum_{w \in W} Q_0(a, b | w), \quad (2)$$

其中 W 是所有窗所组成的集合, $|W|$ 是集合 W 的势。

也可以对每个窗口中的本地质量指数用显著程度加权。假设 $s(a | w)$ 和 $s(b | w)$ 是 a 和 b 在窗口中的显著程度,权值 $\lambda(w)$ 的计算方法如下:

$$\lambda(w) = \frac{s(a | w)}{s(a | w) + s(b | w)}, \quad (3)$$

这样,就可以得到融合质量指数 $Q(a, b, f)$

$$Q(a, b, f) = \frac{1}{|W|} \sum_{w \in W} \{\lambda(w) Q_0(a, f | w) + [1 - \lambda(w)] Q_0(b, f | w)\}, \quad (4)$$

若 $C(w) = \max[s(a | w), s(b | w)]$,使用 $c(w) = C(w) / \sum_{w' \in W} C(w')$ 来代替(4)式中的 $1/|W|$,就得到加权融合质量指数 Q_w 。

根据对(1)式 SSIM 各项物理意义的解释,可以将图像中点 (i, j) 的结构相似标准指数用另一种方式来表达:

$$S_{\text{SIM}}(i, j) = [s(i, j)]^\alpha [l(i, j)]^\beta [c(i, j)]^\gamma, \quad (5)$$

其中 $s(i, j)$ 是结构相似元素, $l(i, j)$ 是亮度相似元素, $c(i, j)$ 是对比度相似元素;常数 α, β, γ 分别表示这 3 个

元素的重要程度。

对比(1)式和(5)式可以发现, α, β, γ 均取1,也就是说3种元素的重要程度被认为相等; $l(x, y) = 2\bar{x}\bar{y}/(\bar{x}^2 + \bar{y}^2)$,即用均值来表示的亮度; $c(x, y) = 2\sigma_x\sigma_y/(\sigma_x^2 + \sigma_y^2)$,即方差 σ_x 来表示 x 的对比度。这些表明SSIM使用的亮度和对比度先验假设来自文献[8]中提出的模型,即

$$I(i, j) \approx N[L(i, j), C(i, j)], \quad (6)$$

也就是说,认为图像信号是一个白色高斯域,像素之间无关系,区域均值为亮度,方差为对比度。

然而这个模型本身是存在一些问题的:首先,它过于简化了背景点的影响;其次,像素间是相互独立的这个假设并不总是成立。正是这些问题导致了SSIM用于图像融合质量评估时,对于不在同一级进行的融合方法进行效果评估时效果不够理想。

3 基于亮度和对比度模型的结构相似度标准

3.1 亮度和对比度模型

存在这样一个现象:当单个像素被显示在屏幕上时,在视觉上它会发生“泄露”^[9],而不只显示一个像素,如图1单个像素点的“泄露”现象所示。



图1 单个像素点的“泄露”现象

因此,对于一个像素而言,它的亮度不仅仅由自身确定,还会受邻域像素的影响。一个像素 (i_0, j_0) 对周围的泄露可以被写成高斯函数的形式^[9]:

$$\Phi(i, j) \propto \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(-i_0)^2 + (j - j_0)^2}{2\sigma^2}\right], \quad (7)$$

其中 σ 的取值因显示器的不同而不同,根据经验一般取值为像素间距 d 的一半, $\sigma = d/2$ 。这样就得到了

$$\Phi(i, j) \propto \frac{1}{\pi d^2} \exp\left[-\frac{2(i - i_0)^2 + 2(j - j_0)^2}{d^2}\right], \quad (8)$$

这个模型更多地考虑了背景像素点和它们之间的相互关系,比旧模型更为贴近人类视觉。

文献[10]提出一种新的对比度模型,并证明该模型可以计算人类视觉敏感度(HVS)的重要参数。该文认为,对比度应该这样计算

$$C(i_0, j_0) = \frac{\max(\text{block}) - \min(\text{block})}{C}, \quad (9)$$

其中block为当前像素点 (i_0, j_0) 为中心的 3×3 区域, C 是整幅图像最大值与最小值的差

3.2 基于新的亮度和对比度模型的结构相似度标准

正如(5)式所示,SSIM包括结构相似、亮度相似和对比度相似3个元素。基于(8)式的亮度模型和(9)式的对比度模型,可以定义新的亮度和对比度相似性,从而定义出新的SSIM。

位于 (i_0, j_0) 处的像素点,它的亮度为自身影响与所有背景点对它影响的总和,也就是说,当前点除去泄露到其他位置的亮度之后在当前位置的亮度,加上背景点对当前点泄露值的总和,即

$$L(i_0, j_0) \propto \sum_{i, j} \Phi(i, j), \quad (10)$$

选择合适的参数,利用(10)式,可以计算当前点的邻域点对其亮度的影响^[9]。图2显示了邻域对于当前点 P ,邻域点的位置。点 P 亮度的47%由自身提供,其他53%来自邻域点的影响。如图2所示,标为A的每个贡献10.7%,标为B的每个贡献2.5%,C、D和E加起来不到0.1%^[9]。

用新的亮度模型来计算亮度相似元素

$$l_n(i, j) = \frac{2L_a(i, j)L_f(i, j)}{L_a^2(i, j) + L_f^2(i, j)}, \quad (11)$$

其中 $l_n(i, j)$ 是位置 (i, j) 处的亮度相似度, a 和 f 分别指代两幅图像, $L_a(i, j)$ 和 $L_f(i, j)$ 分别是两幅图在 (i, j) 位置点的亮度值,通过(10)

E	D	C	D	E
D	B	A	B	D
C	A	P	A	C
D	B	A	B	D
E	D	C	D	E

图2 像素 P 的邻域

Fig. 2 Neighborhood of pixel P

式计算得到。

而对比度相似元素则用下面的式子定义

$$c_n(i, j) = \frac{2C_a(i, j)C_f(i, j)}{C_a^2(i, j) + C_f^2(i, j)}, \quad (12)$$

其中 $c_n(i, j)$ 是位置 (i, j) 处的对比度相似度, a 和 f 分别指代两幅图像, $C_a(i, j)$ 和 $C_f(i, j)$ 分别是两幅图在 (i, j) 位置点的亮度值, 通过(9)式计算得到。

通过这样定义出新的亮度相似度和对比度相似度, 从而得到了基于新亮度和对比度模型的 SSIM 为

$$L_{\text{CSSIM}}(i, j) = [s(i, j)]^\alpha [L_n(i, j)]^\beta [c_n(i, j)]^\gamma. \quad (13)$$

为了消除图像信号不稳定性的影响, 这里也使用类似(2)式滑动窗函数并用(3)式显著度加权, 这样就得到了新的评估度量函数

$$Q_{\text{LCSSIM}}(a, b, f) = \frac{1}{|W|} \sum_{w \in W} \{ \lambda(w) L_{\text{CSSIM}}(a, f | w) + [1 - \lambda(w)] L_{\text{CSSIM}}(b, f | w) \}. \quad (14)$$

4 实验

为了验证新度量标准的有效性, 这里进行了两个实验, 所使用的源图像数据分别是 TNO Human Factors Research Institute 所提供的“UNcamp”序列^[11]和文献[2]所使用的 120 组异源图像。滑动窗和 block 的大小都为 5×5 , 这样选择的理由是: 如果滑动窗太小, 没有包含足够多的背景点对当前点的影响因素, 根据(10)式计算得到的当前点亮度是不完整的; 如果滑动窗过大, 由图 2 和(10)式可以看出, 边缘部分的背景点对当前点的亮度影响几乎可以忽略不计。在实验中假设亮度相似度、对比度相似度和结构相似度的影响式相同, 因此(5)式中的常数 α, β, γ 都取 1。

4.1 UNcamp 序列

本文用 5 种方法对源图像进行融合处理, 分别是均值法, 离散小波变换法(DWT)^[12], 形态学差异金字塔(MDP)^[13]、比率金字塔(RP)^[14]和拉普拉斯金字塔(LP)^[15]。图 3 正是序列中一组源图像及其融合结果。评估这些融合图像效果的方法, 则分别采用文献[4]的 MI, 文献[6]的 Q 和 Q_w , 文献[7]的 Q_m 以及本文提出的 Q_{LCSSIM} 。

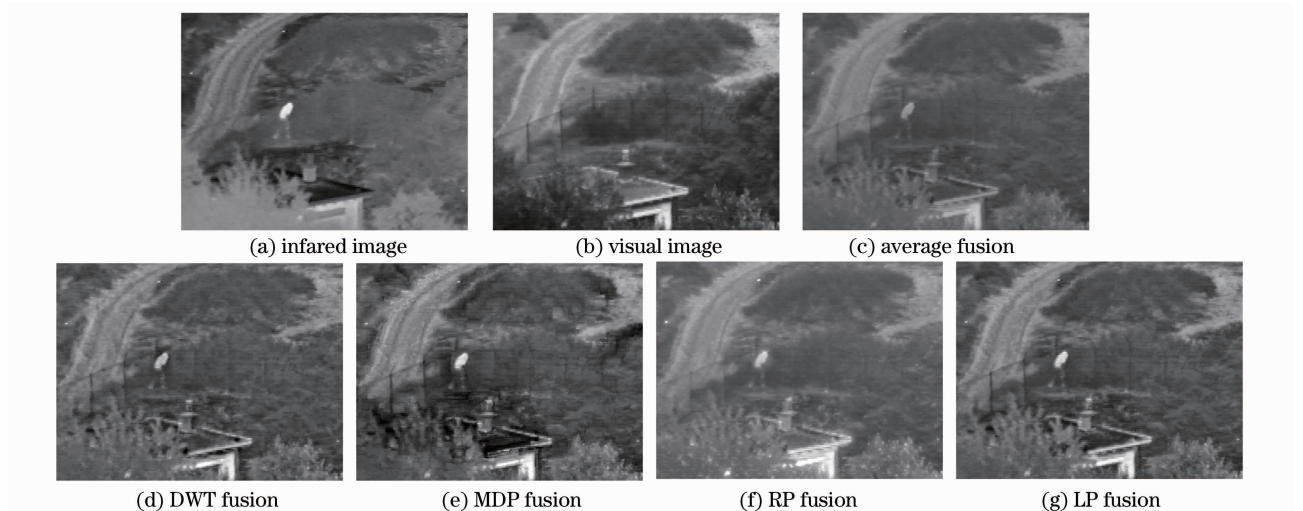


图 3 一组源图像及其融合结果

Fig. 3 Example of source and fused images

从图 3 中能看出, LP、MDP 和 RP 显然都比均值法的结果要好, 保存了两幅源图像中道路痕迹、栏杆和树枝等的细节, 正像表 1 中的 Q_{LCSSIM} 所显示的那样; 而 MI 方法一如既往地给了均值法最高分; Q 将 MDP 和 RP 都评价为低于均值法; 而 Q_w 和 Q_m 也认为 RP 的效果不及均值法。

表 1 对图 3 中各种融合效果的评估结果

Table 1 Evaluation result for fusion image in Fig. 3

	Average	DWT	LP	MDP	RP
MI	1.0778	0.94289	1.0011	1.0157	1.0336
Q	0.62699	0.63664	0.67472	0.57245	0.58719
Q_w	0.64374	0.73769	0.75921	0.67314	0.62989
Q_m	0.81932	0.91139	0.9255	0.93917	0.75645
Q_{LCSSIM}	0.84255	0.90989	0.92182	0.86786	0.88313

整个 32 组图像的评估结果展示在图 4 中。图 4(a)中使用 MI 来评估融合图像效果,均值法的曲线位于所有方法曲线之上,这显然与人的视觉认知结果相悖;在图 4(b)中,Q 认为均值法的表现优于 MDP 和 RP; Q_w 和 Q_m 的评估结果分别见图 4(c)和(d),它们认为均值法效果不及 MDP,但优于 RP。本文提出的新评估标准 Q_{LCSSIM} 在图 4(e)中的曲线走势,将 4 种方法由低到高排序为 LP、DWT、RP、MDP、均值法,直观地验证了 Q_{LCSSIM} 比较符合主观视觉效果的观点。

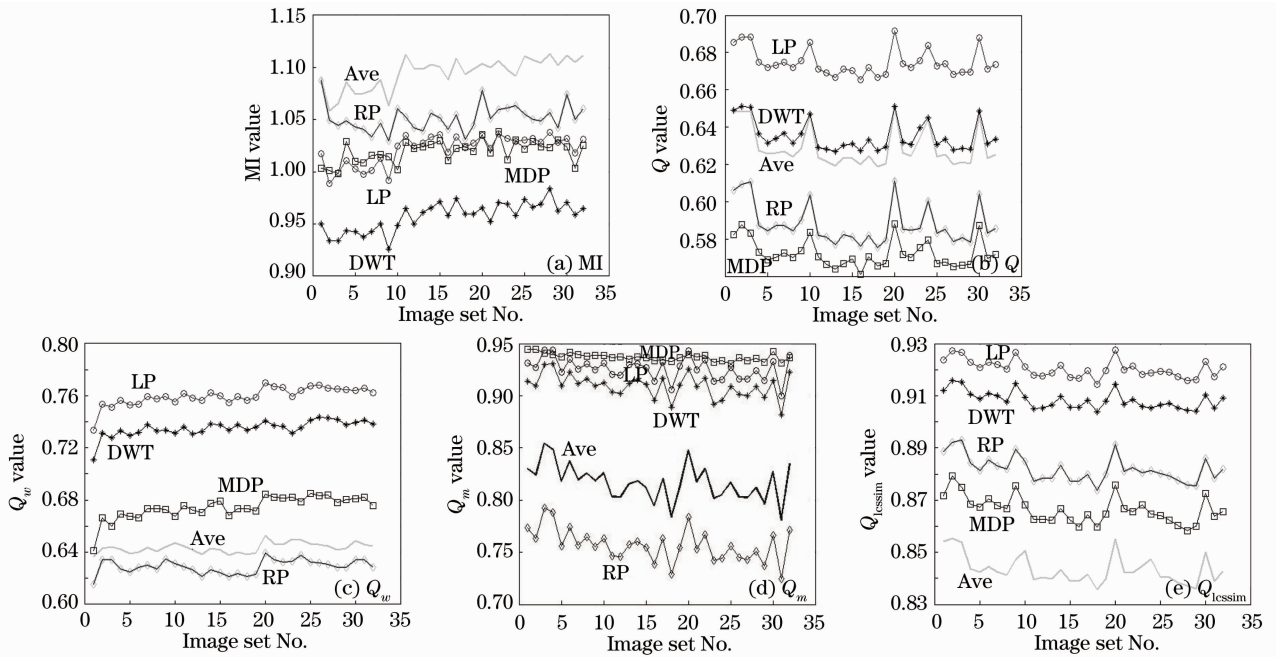


图 4 不同评估方法对全部 32 组源图像融合效果的评估结果

Fig. 4 Evaluation result for all 32 set images

3.4 主观评测标准图像库

图 5 给出了文献[2]所使用的主观评测标准图像库中,一些异源图像实例。在这个实验中,首先也用 5 种方法对 120 组异源图像进行融合,然后对每组融合图像,MI, Q, Q_w 和 Q_{LCSSIM} 每个评估方法会将其按成绩

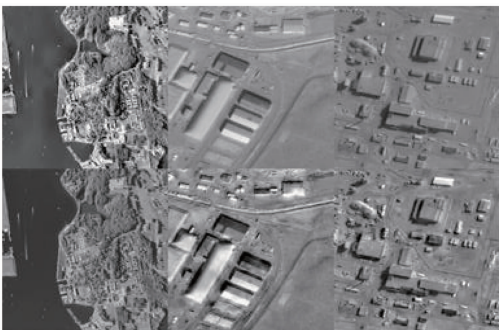


图 5 作为源图像的一些异源图像实例

Fig. 5 Example of source images

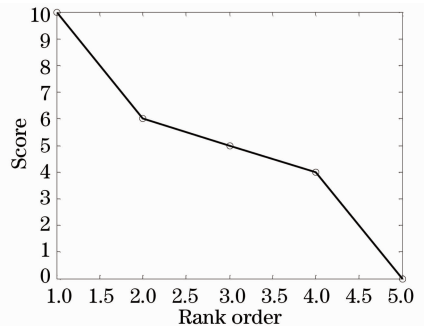


图 6 排名与得分之间的函数关系

Fig. 6 Relation between rank order and scores

的高低排成一个序列。在排名的基础上,为了扩大最优和最差成绩之间的差异,本文对名次用类似于正切函数的函数进行了映射,如图6所示:第一名得10分,第二名6分,第三名5分,第四名4分,第五名0分。

在统计了各个方法的出现在各名次上的情况及其最后的得分情况之后,图7则用饼形图给出了四种评估方法的投票结果。可以看到,MI标准给了均值法最高的分数。 Q 认为MDP和RP方法都劣于均值法,这不符合对这5种融合方法的性能评价。 Q_w 给了MDP高于均值法的成绩,但仍认为RP低于均值法。而新提出的标准 Q_{LCSSIM} 认为,对于这个规模大、涵盖内容广泛、获取方法多样的融合源数据库而言,LP方法最好而均值法最劣。由于均值法是单分辨率融合方法,因此 Q_{LCSSIM} 的打分结果符合图像融合领域中已被证明并被广泛接受的认识:多分辨率融合方法的效果要强于单分辨率。

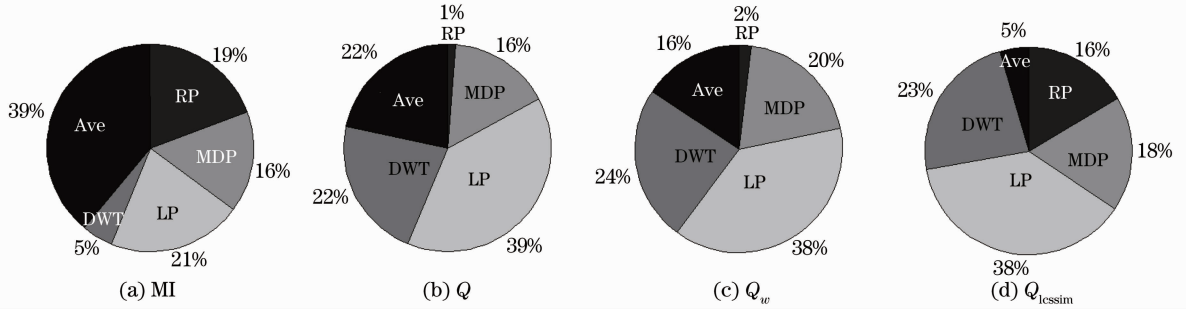


图7 对不同融合方法投票结果的饼形图

Fig. 7 Pie chart of voting results for fusion methods

5 结 论

分析了SSIM的物理模型及其缺点,引入新的亮度和对比度模型,设计出一种新的图像融合质量评估标准。实验结果表明,新引入结构相似度SSIM的亮度和对比度模型,更能够准确反映人的视觉感知与计算机显示硬件的特性,利用它们构造出的新的融合评估标准,也与人的视觉感知标准更为接近。尽管图像融合是一种应用依赖的处理过程,但利用本文提出的评估标准,在评估实施于不同层的融合方法时,还是能够给出合理的结果。

参 考 文 献

- 1 C. Pohl, J. L. Van Genderen. Multisensor image fusion in remote sensing: concepts, methods and applications[J]. *Int. J. Remote Sens.*, 1998, **19**(5): 823~854
- 2 V. Petrovic. Subjective tests for image fusion evaluation and objective metric validation[J]. *Inform. Fusion*, 2007, **8**(2): 208~216
- 3 Z. Liu, D. S. Forsyth, R. Laganiere. A feature-based metric for the quantitative evaluation of pixel-level image fusion[J]. *Comput. Vis. Image Und.*, 2008, **109**(1): 56~68
- 4 Guihong Qu, Dali Zhang, Pingfang Yan. Information measure for performance of image fusion[J]. *Electron. Lett.*, 2002, **38**(7): 313~315
- 5 Z. Wang, A. C. Bovik. A universal image quality index[J]. *IEEE Signal Proc. Lett.*, 2002, **9**(3): 81~84
- 6 Gemma Piella, Henk Heijmans. A new quality metric for image fusion[C]. Proc. IEEE Int. Conf. on Image Processing, 2003. 173~176
- 7 Shanshan Li, Richang Hong, Bo Liu *et al.*. A quality metric based on color similarity for image fusion[J]. *International J. Information Acquisition (IJIA)*, 2008, **5**(3): 235~245
- 8 M. Foracchia, E. Grisan, A. Ruggeri. Luminosity and contrast normalization in retinal images[J]. *Med. Image Anal.*, 2005, **9**(3): 179~190
- 9 Blair MacIntyre, William B. Cowan. A practical approach to calculating luminance contrast on a CRT [J]. *ACM Transactions on Graphics*, 1994, **11**(4): 336~347
- 10 Li Zhonghai, Wang Changhong. Design for filter with human visual sensitivity[J]. *Journal of Image and Graphics*, 2003, **8**(A): 76~79
- 11 A. Toet. Detection of dim point targets in cluttered maritime backgrounds through multisensor image fusion. Targets and

- Backgrounds VIII: Characterization and representation[C]. *SPIE*, 2002, **4718**: 118~129
- 12 Hui Li, B. S. Manjunath, S. K. Mitra. Multisensor image fusion using the wavelet transform[J]. *Graph. Models. Image Process*, 1995, **57**(3): 235~245
- 13 S. Marshall, G. K. Matsopoulos. Morphological data fusion in medical imaging[C]. *Nonlinear Digital Signal Processing*, 1993. *IEEE Winter Workshop*: 6.1_5.1~6.1_5.6
- 14 A. Toet. Image fusion by a ratio of low-pass pyramid[J]. *Pattern Recogn. Lett.*, 1989, **9**(4): 245~253
- 15 P. J. Burt, E. H. Adelson. The laplacian pyramid as a compact image code[J]. *IEEE trans. Commun.*, 1983, **31**(4): 532~540