# An Attention-Based Decision Fusion Scheme for Multimedia Information Retrieval

Xian-Sheng Hua, Hong-Jiang Zhang

Microsoft Research Asia
5F Sigma Center, 49 Zhichun Road, Beijing 100080, P.R.China
{xshua, hjzhang}@microsoft.com

**Abstract.** In this paper, we proposed a novel decision fusion scheme based on the psychological observations on human beings' visual and aural attention characteristics, which combines a set of decisions obtained from different data sources or features to generate better decision result. Based on studying of the "heterogeneity" and "monotonicity" properties of certain types of decision fusion issues, a set of so-called *Attention Fusion Functions* are devised, which are able to obtain more reasonable fusion results than typical fusion schemes. Preliminary experiment on image retrieval shows the effectiveness of the proposed fusion scheme.

## 1 Introduction

Generally information fusion is about summarizing information in an embodiment of multiple sources and typically categorized into three classes, data fusion, feature fusion and decision fusion [1]. Considerable works on this topic have been reported in literatures [1]-[8]. In the field of multimedia content analysis, indexing and retrieval, numerous theory or application issues can be classified into one or more of these three classes. Among these, decision fusion is aiming at obtaining better or optimal decision result by appropriately combining a set of decision results, whether they are hard decisions (i.e., true of false) or soft ones (confidence values), from different sensors, feature sets, and so on. A good decision fusion scheme is expected to sufficiently utilize the information provided by the set of to-be-fused decision results, while suppress the noises in them at the same time.

In this paper, we will propose a new decision fusion scheme based on the psychological observations on human being's visual and aural attention characteristics, which combines a set of decisions obtained from different data sources or features to generate better decision result. "Attention" is a neurobiological conception. It means the concentration of mental powers upon an object by close or careful observing or listening, which is the ability or power to concentrate mentally [9]. Generally, people will concentrate their attention upon the circumstances when there is something (event, object, etc.) can be apperceived. And the "degree" that people will concentrate their attention on the event or object is proportional to the "quantity" of the information (or we may call it *attention index* or *attention value* [9]) it provides, say, the strength of a sound, the speed of a motion, the size of an object, and so on. But what will happen when two or more sources of such information are provided in the circumstance? Are people will concentrate their attention in the degree of the "sum" of all of the attention indices, or the average of them, or the maximum of them? Generally linear combination (LC) of the attention indices of different attention components (e.g., motion, color, audio, etc.) is a simple but effective scheme. However, this kind of linear combination is not reflecting all the information that the attention indices of the attention components contained [10]. In this paper, we proposed a so-called *Attention Fusion Function* (AFF) to model the above issues. And, this fusion scheme can be used to fuse a set of decisions with similar characteristics (we may call it "attention properties", to be detailed in Section 2). It should be mentioned that the proposed AFF fusion scheme is not a general fusion scheme suited for solving general decision fusion issues, but especially applicable for decision issues having the "attention properties".

The rest of the paper is organized as follows. Section 2 presents the attention-based decision fusion scheme in detail. Preliminary experiments on applying the fusion scheme on image retrieval are introduced in Section 3, followed by conclusion remarks in Section 4.

## 2 Attention Fusion

We denote the set of to-be-fused decision results (normalized to interval [0,1]) as a decision vector $\vec{x} = (x_1, x_2, \cdots, x_n)$, where $0 \le x_i \le 1, 1 \le i \le n$. A general fusion function is denoted as $f(\vec{x})$ or $f(x_1, x_2, \cdots, x_n)$. It should be noted that the following analyses are based on the assumption that the to-be-fused decisions have the "attention properties" mentioned above.

### 2.1 Two-Dimensional Case

Firstly let's consider a simple case in which we only have two decisions to fuse, i.e., $n = 2$. Weights for the decisions are also not taken into account at this stage, i.e., all the weights are equal to $1/n$. Let's see two decision vectors, (0.8, 0) and (0.4, 0.4), if linear combination is applied, $f(0.8, 0)$ will have the same value as $f(0.4, 0.4)$. However, this result does not coincide with the real case. Actually, the first decision vector is more "attractive" (if we regard the decisions as attention indices of different attention components), as one attention component with high attention index will "attract" people's attention greatly. Accordingly, it will be better if the fusion function satisfies the following inequality,

$$f(x_1, x_2) < f(x_1 + \varepsilon, x_2 - \varepsilon) , \quad \textbf{(1)}$$

where $0 < \varepsilon \le x_2 \le x_1$. In contrast, if we use linear combination, $f(x_1, x_2)$ is equal to $f(x_1+\varepsilon, x_2-\varepsilon)$, thus it does not satisfy this property. For convenience, we name this property "heterogeneity". On the other hand, it is obvious that the fusion function should satisfy another property, monotonicity, i.e.,

$$f(x_1, x_2) < f(x_1 + \varepsilon, x_2) , \quad \textbf{(2)}$$

where $\varepsilon > 0$. Maximum function (MAX) satisfies (1), while does not strictly satisfy this monotonicity property. When the strict inequality signs in equality (1) and (2) are replaced by non-strict inequality signs, the fusion function could be,

$$AFF^{(0)}(x_1, x_2) = \frac{1}{2}\left[(x_1 + x_2) + |x_1 - x_2|\right] \cdot \quad \textbf{(3)}$$

Obviously, this function is obtained just by adding a correction, the difference between the two decisions, to the linear combination fusion result. In order to strictly satisfy the two inequalities simultaneously, we have the following theorem.

**Theorem 1** (*2-Dimensional AFF without Weights*): The following function

$$AFF_2^{(\gamma)}(x_1, x_2) = \frac{1}{2}\left[(x_1 + x_2) + \frac{1}{1+\gamma}|x_1 - x_2|\right] \quad \textbf{(4)}$$

satisfies inequality (1) and (2), where $\gamma > 0$ is a constant. □

For the above mentioned example, we got $AFF_2^{(0.2)}(0.8, 0) = 0.733$, while $AFF_2^{(0.2)}(0.4, 0.4) = 0.4$, which obviously indicates the first one is more "attractive". Figure 1 shows the comparison of *LC*, *MAX*, $AFF^{(0)}$ and $AFF_2^{(\gamma)}$. Actually if $x_1 = x_2$, we will have $MC = MAX = AFF_2^{(\gamma)} = AFF^{(0)}$, while when $x_1 \neq x_2$, we have

$$LC(x_1, x_2) < AFF_2^{(\gamma)}(x_1, x_2) < AFF_2^{(0)}(x_1, x_2) = MAX(x_1, x_2)\ . \quad \textbf{(5)}$$

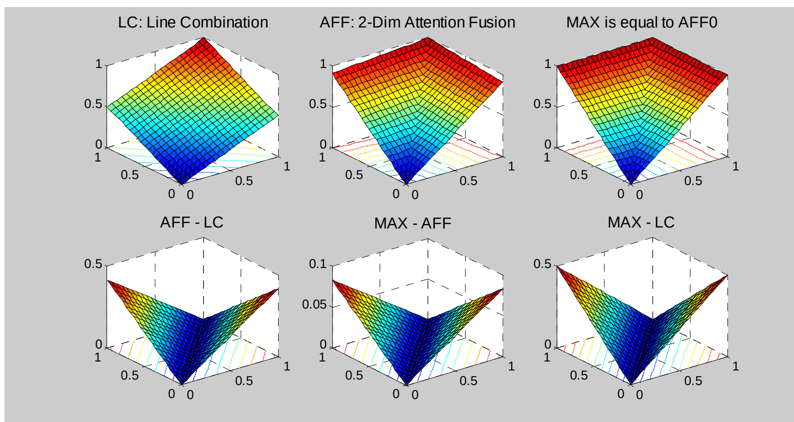



**Fig. 1.** Comparison of *LC*, *MAX*, $AFF^{(0)}$ (denoted by "AFF0" in the figure), and $AFF_2^{(\gamma)}$ (denoted by "AFF"). The three sub-figures in the second line show the differences between $AFF_2^{(\gamma)}$ and LC, MAX and $AFF_2^{(\gamma)}$, MAX and LC, respectively.

The parameter $\gamma > 0$ is a predefined constant, which controls the amount of differences between the left sides and right sides of inequalities (1) and (2) when $x_1$, $x_2$ and $\varepsilon$ are fixed. The greater the parameter $\gamma$ is, the smaller the differences are. To be exact, $(1/\gamma)$ represents the effectiveness of one decision component in the overall decision. For example, $f(0.5, 0.7) - f(0.6, 0.6)$ is equal to 0.091 and 0.067 when $\gamma$ is equal to 0.1 and 0.5, respectively. The smaller the parameter $\gamma$ is, the more greatly that one decision component with high index (or confidence) will affect (increase) the overall decision index. In Table 1, we list some examples to show the differences between attention fusion function and LC (i.e., average) under different parameters.

**Table 1.** Differences between attention fusion function and averaging

| $x_1$ | $x_2$ | $\gamma$ | $AFF_2^{(\gamma)}$ | Average | Difference | $\gamma$ | $AFF_2^{(\gamma)}$ | Average | Difference |
|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.8 | 0.2 | 0.74 | 0.45 | 0.29 | 0.5 | 0.68 | 0.45 | 0.23 |
| 0.2 | 0.8 | | 0.75 | 0.50 | 0.25 | | 0.70 | 0.50 | 0.20 |
| 0.5 | 0.5 | | 0.50 | 0.50 | 0.00 | | 0.50 | 0.50 | 0.00 |
| 0.0 | 1.0 | | 0.91 | 0.50 | 0.41 | | 0.83 | 0.50 | 0.33 |

### 2.2 *n*-Dimensional Case

For *n*-dimensional case we have **Theorem 2**.

**Theorem 2** (*n-Dimensional AFF without Weights*): The following function

$$AFF_n^{(\gamma)}(\vec{x}) = E(\vec{x}) + \frac{1}{2(n-1)+n\gamma}\sum_{k=1}^{n}\left|x_k - E(\vec{x})\right| , \quad \textbf{(6)}$$

where $\gamma > 0$ is a predefined constant, and $E(\vec{x})$ is the average value of decision components in vector $\vec{x}$, satisfies the inequality

$$AFF_n^{(\gamma)}(x_1, ..., x_i, ..., x_n) < AFF_n^{(\gamma)}(x_1, ..., x_i + \varepsilon, ..., x_n) , \quad \textbf{(7)}$$

where $1 \le i \le n$, $\varepsilon \ge 0$, and inequality

$$AFF_n^{(\gamma)}(x_1, ..., x_i, ..., x_j, ..., x_n) \le AFF_n^{(\gamma)}(x_1, ..., x_i + \varepsilon, ..., x_j - \varepsilon, ..., x_n) , \quad \textbf{(8)}$$

where $1 \le i < j \le n$, $x_i \ge x_j \ge \varepsilon > 0$. When

$$x_j - \varepsilon \le E(\vec{x}) \le x_i + \varepsilon , \quad \textbf{(9)}$$

the inequality sign in (8) will be strictly satisfied. □

When weights are taken into consideration, we have

**Theorem 3** (*n-Dim AFF with Weights*): The following function

$$AFFW_n^{(\gamma)}(\vec{x}) = \frac{\vec{w}\cdot\vec{x} + \frac{1}{2(n-1)+n\gamma}\sum_{k=1}^{n}\left|nw_k x_k - \vec{w}\cdot\vec{x}\right|}{W} \quad \textbf{(10)}$$

satisfies inequality

$$AFFW_n^{(\gamma)}(x_1, ..., x_i, ..., x_n) < AFFW_n^{(\gamma)}(x_1, ..., x_i + \varepsilon, ..., x_n) \quad \textbf{(11)}$$

where $1 \le i \le n$, $\varepsilon \ge 0$, and inequality

$$AFFW_n^{(\gamma)}(x_1, ..., x_i, ..., x_j, ..., x_n) \le AFFW_n^{(\gamma)}(x_1, ..., x_i + w_j\varepsilon, ..., x_j - w_i\varepsilon, ..., x_n) \quad \textbf{(12)}$$

where $1 \le i < j \le n$, $w_i x_i \ge w_j x_j \ge w_i w_j \varepsilon > 0$. When

$$nw_j x_j - nw_i w_j \varepsilon \le \vec{w}\cdot\vec{x} \le nw_i x_i + nw_i w_j \varepsilon , \quad \textbf{(13)}$$

the inequality sign in (12) will be strictly satisfied. In the above expressions, $\gamma > 0$ is a predefine constant, and

$$W = W(\vec{w}) = 1 + \frac{1}{2(n-1)+n\gamma}\sum_{k=1}^{n}\left|1 - nw_k\right| , \quad \textbf{(14)}$$

and $\vec{w} = (w_1, \cdots, w_n)$ is the weight vector satisfying $w_k \ge 0$, $1 \le k \le n$, and $\sum_{k=1}^{n} w_k = 1$. □

Actually, *W* is equal to the maximum value of the numerator of the right side of equation (10). Therefore, we have

$$0 \le AFFW_n^{(\gamma)}(\vec{x}) \le 1 \ . \quad \textbf{(15)}$$

When all weights are equal, obviously we have $W = 1$ and

$$AFFW_n^{(\gamma)}(\vec{x}) = AFF_n^{(\gamma)}(\vec{x}) \ . \quad \textbf{(16)}$$

Figure 2 shows the comparison of *LC*, $AFF_2^{(\gamma)}$ and $AFFW_2^{(\gamma)}$, when $w_1 = 0.6$ and $w_2 = 0.4$.

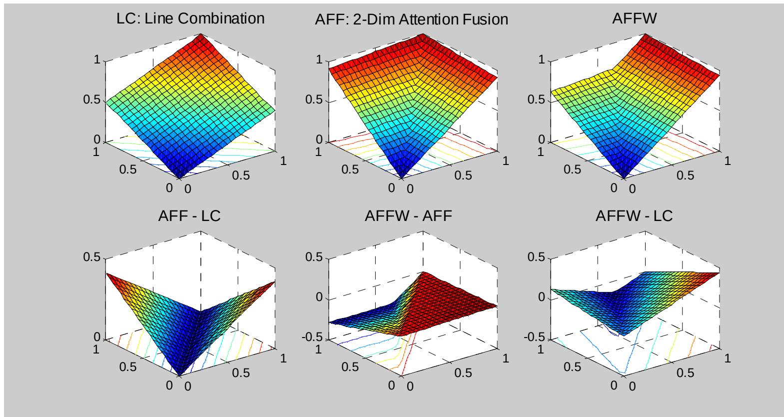



**Fig. 2.** Comparison of *LC*, $AFF_2^{(\gamma)}$ (denoted by "AFF" in the figure) and $AFFW_2^{(\gamma)}$ (denoted by AFFW). The figures in the second line show the differences between $AFF_2^{(\gamma)}$ and LC, $AFFW_2^{(\gamma)}$ and $AFF_2^{(\gamma)}$, $AFFW_2^{(\gamma)}$ and *LC*, respectively.

### 2.3 Further Improvement

It is observed that, for *n*-dimensional AFF, some fusion results are still not "fair" enough. For example, when *n* = 10, $f(0.1, 0.1, \cdots, 0.1) = 0.1$ while $f(1, 0, \cdots, 0) = 0.19$ ($\gamma = 0.2$). Intuitively, $f(1, 0, \cdots, 0)$ should have relatively larger value. Actually, according to (6) and (8), it is easy to prove that

$$LC(\bar{x}) \le AFF_n^{(\gamma)}(\bar{x}) \le AFF_n^{(0)}(\bar{x}) \le \frac{2MAX(\bar{x})}{n} . \quad \textbf{(17)}$$

Therefore, when *n* is large, $AFF_n^{(\gamma)}(\bar{x})$ is much less than $MAX(\bar{x})$. To solve this issue, another constraint is required, which is defined by

$$f(\bar{x}) \ge \alpha \cdot MAX(\bar{x}), \quad \textbf{(18)}$$

where $\alpha$ is a predefined constant.

**Theorem 4** (*Improved n-Dimensional AFF with Weights*): The following function

$$IAFFW_n^{(\gamma)}(\bar{x}) = \frac{\alpha \cdot \max_{1\le k\le n}\{nw_k x_k\} + \beta \cdot W \cdot AFFW_n^{(\gamma)}(\bar{x})}{W^*} \quad \textbf{(19)}$$

satisfies inequality (11), (12) (function name should be replaced by IAFFW) under the corresponding constraints, as well as satisfies

$$IAFFW_n^{(\gamma)}(\bar{x}) \ge \frac{\alpha}{W^*} \cdot \max_{1\le k\le n}\{nw_k x_k\} \quad \textbf{(20)}$$

$$0 \le IAFFW_n^{(\gamma)}(\bar{x}) \le 1 \quad \textbf{(21)}$$

where $0 < \alpha, \beta < 1$, $\alpha + \beta = 1$, $W$ is defined in equation (14), and

$$W^* = \alpha \cdot \max_{1\le k\le n}\{nw_k x_k\} + \beta \cdot W \quad \cdot \quad \textbf{(22)}$$

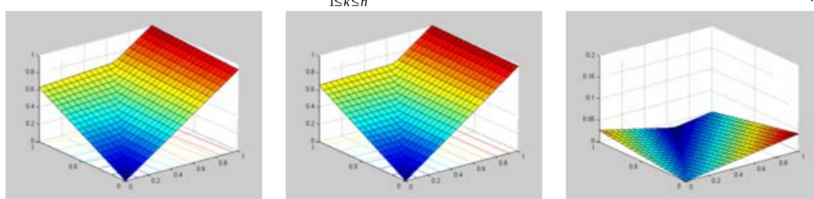
When we have (13) or

$$w_i x_i = \max_{1\le k\le n}\{w_k x_k\} \quad \textbf{(23)}$$

the inequality sign in (12) will be strictly satisfied. □

Proof of **Theorem** 4 can be found in the Appendix, which also indicates **Theorem** 1, 2 and 3 are also correct. Figure 3 shows the comparison of $AFFW_2^{(0.2)}(\vec{x})$ and $IAFFW_2^{(0.2)}(\vec{x})$. In the above example, when we set $\alpha = 0.7, \beta = 0.3, \gamma = 0.2$, and all weights are equal to 0.1, we have better results as $f(0.1, 0.1, \cdots, 0.1) = 0.1$ and $f(1, 0, \cdots, 0) = 0.757$. Obviously, if all weights are equal, we have $W^* = W = 1$, and the improved attention fusion function without weight is

$$IAFF_n^{(\gamma)}(\vec{x}) = \alpha \cdot \max_{1 \le k \le n}\{x_k\} + \beta \cdot AFF_n^{(\gamma)}(\vec{x}) \ . \qquad \textbf{(24)}$$



(a) $AFFW_2^{(0.2)}(\vec{x})$ (b) $IAFFW_2^{(0.2)}(\vec{x})$ (c) Difference between (b) and (a)

**Fig. 2.** Comparison of $IAFFW_2^{(0.2)}(\vec{x})$ and $AFFW_2^{(0.2)}(\vec{x})$ ($\alpha = 0.7, \beta = 0.3$, $w_1 = 0.6, w_2 = 0.4$)

## 3. Preliminary Experiments

Content-based image retrieval is taken as an example to demonstrate the advantage of the novel fusion scheme. A database containing 10,000 images excerpted from Coral Draw image database is used, which has been categorized into 79 classes (such as tiger, beach, building, etc.) according to their content. In the experiments, 20% images randomly selected from the database are applied as query samples one-by-one, while the images in the same class as the query image are considered as positive samples.

Six sets of features, including color histogram, color moment, wavelet, block wavelet, correlagram, blocked correlagram [11], are employed to calculate the similarity between two images using L1 distance measure. Each set of features will produce a similarity measure or decision for any image in the database compared with the query image. Table 2 shows the performances (recall rate for the first 100 query results and precision rate for the first 10, 20, 30 and 100 query results) when using average fusion, maximal fusion and attention fusion (IAFFW), respectively. From the numbers in Table 2, it can be seen that Attention Fusion produces relatively better results.

**Table 2.** Comparison of different decision fusion schemes. R100 is the recall rate for the first 100 query results, while P10, P20, and P30 are the precision rates for the first 10, 20 and 30 query results, respectively. And the parameters for IAFFW are set as, $\gamma = 0.2$, $\alpha = 0.7$, $\beta = 0.3$, $w_1 = w_2 = \cdots = w_6 = 1/6$

| Fusion Schemes | R100 | P10 | P20 | P30 | P100 |
|---|---|---|---|---|---|
| Average Fusion | 0.121 | 0.315 | 0.244 | 0.212 | 0.138 |
| Maximal Fusion | 0.125 | 0.321 | 0.252 | 0.230 | 0.141 |
| IAFFW | 0.130 | 0.325 | 0.265 | 0.243 | 0.153 |

A better Evaluation of the proposed attention fusion scheme can be obtained by applying the scheme on the decisions produced from different types of features, such as color, shape and so on, as well as by testing on applications in which the decisions are obtained from multiple modalities. This is our future work.

## 4. Conclusion and Discussion

In this paper, we have proposed a novel decision fusion scheme, Attention Fusion, which combines a set of decisions generated from different data sources or features to obtain better decision result. The fusion scheme is based on the two properties, monotonicity and heterogeneity, of the to-be-fused decision set, which come from psychological observations and assumptions of human beings' visual and aural attentions. The proposed fusion scheme can be used to obtain better decision result in the case of these two properties are satisfied. The future work would be to test the scheme on more and wider applications. In addition, how to automatically determine the best parameters is still unsolved. Another future work would be to extend the idea of constructing attention functions to other types of decision fusion issues with different properties.

## References


1. Dasarathy, B.V.: Decision Fusion. IEEE Computer Society, August 1993
2. Samarasooriya, V.N.S., et al: A Fuzzy Modeling Approach to Decision Fusion Under Uncertainty. IEEE /SICE/RSJ Intl. Conf. on Multisensor Fusion and Integration of Intelligent Systems (1996), 788-795.
3. Chen, B., Varshney, P.K.: A Bayesian Sampling Approach to Decision Fusion Using Hierarchical Models. IEEE Trans. on Signal Processing (2002), Vol. 50, No. 8
4. Li, X.R., Zhu, Y., Wang, J., Han, C.: Optimal Linear Estimation Fusion – Part I: Unified Fusion Rules. IEEE Trans. on Information Theory (2003), Vol. 49, No. 9
5. Woods, K., *et al*: Combination of Multiple Classifiers Using Local Accuracy Estimates, IEEE Transactions on Pattern Analysis and Machine Intelligence (1997), April
6. Petrakos, M., *et al*: The Effect of Classifier Agreement on the Accuracy of the Combined Classifier in Decision Level Fusion, IEEE Transactions on Geoscience and Remote Sensing (2001), Nov.
7. Ji, C., Ma, S.: Combinations of Weak Classifiers. IEEE Transactions on Neural Networks (1997), Jan.
8. Xu, L., *et al*: Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition, IEEE Transactions on Systems, Man and Cybernetics(1992), May/June
9. Ma, Y.F., Lu, L., Zhang, H.J., Li, M.J.: An Attention Model for Video Summarization. ACM Multimedia (2002), Juan-les-Pins, France, December
10. Ma, Y.F., et al: User Attention Model based Video Summarization. To appear in IEEE Transactions on Multimedia Journal.
11. Veltkamp, R. C., Tanase, M.: Content-based image retrieval systems: a survey. March 2001, http://www.aa-lab.cs.uu.nl/cbirsurvey/cbir-survey/.


## Appendix

**Proof of Theorem 4**: Without losing generality, we let $i = 1$, and $j = 2$.

**(a) Proof of inequality (12):**

$$\begin{aligned} &IAFFW_n^{(\gamma)}(\vec{x}') = IAFFW_n^{(\gamma)}(x_1', x_2', ..., x_n') = IAFFW_n^{(\gamma)}(x_1 + w_2\varepsilon, x_2 - w_1\varepsilon, x_3, ..., x_n) \\ &= \frac{\alpha \cdot \max\{nw_1x_1 + nw_1w_2\varepsilon, nw_2x_2 - nw_1w_2\varepsilon, nw_3x_3, ..., nw_nx_n\}}{W^*} + \frac{\beta \cdot W \cdot AFFW_n^{(\gamma)}(\vec{x}')}{W^*} \end{aligned} . \quad \textbf{(25)}$$

As $w_1x_1 \geq w_2x_2 \geq w_1w_2\varepsilon$, we have

$$\max\{nw_1x_1 + nw_1w_2\varepsilon, nw_2x_2 - nw_1w_2\varepsilon, nw_3x_3, ..., nw_nx_n\} \geq \max_{1 \leq k \leq n}\{nw_kx_k\} \cdot \quad \textbf{(26)}$$

On the other hand,

$$W\cdot AFFW_n^{(\gamma)}(\vec{x})=\frac{1}{n}\sum_{k=1}^{n} nw_k x_k+\frac{1}{2(n-1)+n\gamma}\sum_{k=1}^{n}\left|nw_k x_k-\frac{1}{n}\sum_{k=1}^{n} nw_k x_k\right| \quad . \tag{27}$$

Let

$$y_k=nw_k x_k,\; y_k'=nw_k x_k',\; \varepsilon'=nw_1w_2\varepsilon \quad . \tag{28}$$

Then we have

$$W\cdot AFFW_n^{(\gamma)}(\vec{x})=E(\vec{y})+\frac{1}{2(n-1)+n\gamma}\sum_{k=1}^{n}\left|y_k-E(\vec{y})\right| \tag{29}$$

and

$$W\cdot AFFW_n^{(\gamma)}(\vec{x}')=E(\vec{y})+\frac{1}{2(n-1)+n\gamma}\sum_{k=1}^{n}\left|y_k'-E(\vec{y})\right| \quad . \tag{30}$$

As

$$y_1=nw_1x_1\ge nw_2x_2=y_2 \quad , \tag{31}$$

it is easy to prove (by removing the absolute signs under certain conditions) that

$$\left|y_1'-E(\vec{y})\right|+\left|y_2'-E(\vec{y})\right|=\left|y_1+\varepsilon'-E(\vec{y})\right|+\left|y_2-\varepsilon'-E(\vec{y})\right|\ge\left|y_1-E(\vec{y})\right|+\left|y_2-E(\vec{y})\right| . \tag{32}$$

From (29), (30) and (32), we have

$$W\cdot AFFW_n^{(\gamma)}(\vec{x}')\ge W\cdot AFFW_n^{(\gamma)}(\vec{x}) \quad . \tag{33}$$

Consequently, from (25), (26) and (33) we have

$$IAFFW_n^{(\gamma)}(\vec{x}')\ge IAFFW_n^{(\gamma)}(\vec{x}) \quad . \tag{34}$$

When (13) is satisfied, the inequality sign in (32) and (33) will be strictly satisfied. While if (23) is satisfied, the inequality sign in (26) will be strictly satisfied. Therefore, when (13) or (23) is satisfied, the inequality sign in (34) will strictly satisfied.

**(b) Proof of inequality (11):** Similar to (a), we only need to verify

$$W\cdot AFFW_n^{(\gamma)}(\vec{x}')>W\cdot AFFW_n^{(\gamma)}(\vec{x}) \tag{35}$$

where

$$\vec{x}'=(x_1+\varepsilon,x_2,\cdots,x_n) \quad . \tag{36}$$

We use (28) to define $\vec{y}=(y_1,y_2,\cdots,y_n)$ and $\vec{y}'=(y_1',y_2',\cdots,y_n')$, we then have

$$\begin{aligned}&(2(n-1)+n\gamma)E(\vec{y}')+\sum_{k=1}^{n}\left|y_k'-E(\vec{y}')\right|\\&=(2(n-1)+n\gamma)\left(E(\vec{y})+\frac{w_1\varepsilon}{n}\right)+\\&\quad\left|y_1+w_1\varepsilon-\left(E(\vec{y})+\frac{w_1\varepsilon}{n}\right)\right|+\sum_{k=2}^{n}\left|y_k-\left(E(\vec{y})+\frac{w_1\varepsilon}{n}\right)\right|\end{aligned} \tag{37}$$

$$\begin{aligned}&=(2(n-1)+n\gamma)E(\vec{y})+\left[\left|y_1+w_1\varepsilon-\left(E(\vec{y})+\frac{w_1\varepsilon}{n}\right)\right|+\frac{n-1}{n}w_1\varepsilon\right]\\&\quad+\sum_{k=2}^{n}\left[\left|y_k-\left(E(\vec{y})+\frac{w_1\varepsilon}{n}\right)\right|+\frac{w_1\varepsilon}{n}\right]+\gamma w_1\varepsilon\end{aligned} \tag{38}$$

$$\ge(2(n-1)+n\gamma)E(\vec{y})+\left|y_1-E(\vec{y})\right|+\sum_{k=2}^{n}\left|y_k-E(\vec{y})\right|+\gamma w_1\varepsilon \tag{39}$$

$$>(2(n-1)+n\gamma)E(\vec{y})+\sum_{k=1}^{n}\left|y_k-E(\vec{y})\right| \quad . \tag{40}$$

Thus (35) is proved, and consequently inequality (11) is proved.

**(c) Proofs of inequality (20) and (21) are obvious, so they are omitted here.** ☐