LEARNING INTERACTION MEASURE WITH RELEVANCE FEEDBACK IN IMAGE RETRIEVAL

NGO TRUONG GIANG[1], NGO QUOC TAO[2], NGUYEN DUC DUNG[2], NGO HOANG HUY[2]

[1]*Faculty of Information Technology, HaiPhong Private University;*
[2]*Institute of Information Technology, Vietnam Academy of Science and Technology;*
giangnt@hpu.edu.vn,{nqtao,nddung,nhhuy}@ioit.ac.vn

Abstract. Relevance feedback is an effective approach to bridge the gap between low-level feature extraction and high-level semantic concept in content-based image retrieval (CBIR). In this paper, the authors further improve the use of users feedback with multi-feature query and the Choquet integral. Taking into account the interaction among feature sets, feedback information is used to adjust the feature's relevance weights that are considered as the fuzzy density values in the Choquet integral to define the overall similarity measure between two images. The feature weight adjustment and integration aims at minimizing the difference between users desire and outcome of the retrieval system. Experimental results on several benchmark datasets have shown the effectiveness of the proposed method in improving the quality of CBIR systems.

Keywords. Content-based image retrieval, relevance feedback, linear programming, fuzzy measures, Choquet integral.

1. INTRODUCTION

Image retrieval has become an important research topic in multimedia applications and attracted attention of many researchers in last decades. Basically, there are two main approaches in building an image retrieval system: text-based and content-based [22, 11, 17]. In text-based image retrieval systems, the users' queries are composed by key-words, which describe image content. The system retrieves images using image labels which are annotated manually. However, the difficulties in annotating a massive number of images and the avoiding subjectively labeling make this framework impractical in many situations.

In contrast to the keywords description, content-based image retrieval (CBIR) uses visual features such as colors, textures, patterns and shapes to describe image content. Its advantage over the keyword based image retrieval lies in the fact that the features are automatically extracted and the image's content. However, the most challenging in CBIR is the semantic gap between human understanding (high level concepts) and machine understanding (low-level features) [22]. Many algorithms have been proposed to describe low level features of image and compute the similarity between them [5, 31, 32, 14, 37, 8], the semantic gap has not yet been satisfactorily resolved and the performance of CBIR is still far from the expectations of the user.

One possible solution to narrow down the semantic gap is integrating human interaction into CBIR systems, which is popularly known as Relevance Feedback (RF) [34, 30]. Basically, RF is a supervised learning technique that aims to improve the retrieval performance by learning the subjective

perception of users about the similarity between the images. According to the method, the system needs to be run through several iterations. For each iteration, the system retrieves a series of images that are ordered according to a pre-defined similarity measure, and it requires user's interaction to mark the relevant and non-relevant retrievals [2]. These data are then used to update the reformulated query vector for helping next query results that are being nearer the user's requirement [21]. Therefore, the query model used in the system should be designed to be flexibly enough so that it can be dynamically adjusted to approximate a variety of similarity adjustments encountered in query process.

Multi-feature query paradigm used in CBIR system aims to mine the complementary information of multiple features for achieving strong generalization. In this paradigm, each image is represented by a multi-dimensional feature vector, and each element of the feature vector is assigned the different weights which based on the feedback of users. These weights are then used for aggregating to produce a definitive result that may consist of a composite similarity or score value [3]. Many methods for constructing an aggregation of similarities have been developed. Rui et al. [26] proposed an interactive retrieval approach in which, each image is represented by vectors with the corresponding weights that represent the relative importance of each feature within a vector as well as their importance across the entire data set. These weights are dynamically updated based on user's subjective perception so as to place more weight on relevant elements and less on irrelevant ones. Iqbal et al. [19] used a technique based on Gaussian normalization to process distances of color, texture and structure measures. Then, a global measure is defined as a weighted linear aggregation of the normalized distances. In [10], a set of weights for each descriptor is derived based on genetic algorithm, then they are used to combine image database descriptors. Knowles et. al. [35] have defined a global measure as an optimal linear aggregation of partial similarity functions by using the multi-objective optimization technique by a Pareto Archive Evolution Strategy [20]. In [3], a new similarity measure is designed based on assuming statistical independence, which combines the results obtained with each independent function. This approach is expanded in [4] by generating a straightforward rule that can be easily implemented. The negative feedback is used to identify the context of positive example selection as well as disambiguate the concept being searched. The key idea behind above methods is to change the weight of each image feature. However, the mentioned methods consider the features in terms of independent elements. And the additive operators as the weighted sum are often used to produce an overall similarity measure. Therefore, they are not flexible enough to model queries when several of features may be important to the user only if some other feature is present or absent.

A solution to overcome the above problems is to use the Sugeno measure [25] with the Choquet integral (CI) [24]. In [13], the Sugeno measure is used to define the importance of each component feature and the Choquet integral is used to estimate the similarity between images. However, a limitation of Sugeno measure is considered all pairs of features interact the same way which is not appropriate in the context of retrieval using multiple feature spaces.

In this paper, the Choquet integral is used as a similarity measure, and a effective algorithm is proposed to learn the relevant weight of the features based on the relevance feedback. The user's feedback information is modeled by a fuzzy set used to learn relevant weight of the features. These weights are then used as the fuzzy density functions in Choquet integral to aggregate individual similarity measures in order to produce overall similarity measure. The learning relevant weight of the features is stated in the least absolute deviation sense that is solved by linear programming techniques. This allows us to estimate the relevant weights not only to each feature, but also to

their aggregations that represents query image more closely to user's target. Experimental results on several benchmark datasets show that the proposed method effectively improve the quality of the CBIR system. The paper is organized as follows. The brief of discrete Choquet integrals and some related concepts are presented in the Section 2. Section 3 presents detailed descriptions of the proposed method for combining the similarity measures based on relevance feedback. The empirical results are presented in section 4. Finally, some conclusions and future research directions are drawn.

2. FUZZY MEASURES AND CHOQUET INTEGRAL

In this section, the basic concepts of the Choquet integral is recalled. This is a tool for data fusion, which is documented, e.g. in [28, 16, 15, 25].

2.1. Fuzzy measures

Definition 1. *[28] Let X be a set, $X \neq \emptyset$. A fuzzy measure (also called capacity) on X is a set function $g : 2^X \to [0,1]$ satisfying the following conditions:*

- Boundary conditions: $g(\emptyset) = 0, g(X) = 1$;

- Monotonicity: $g(A) \leq g(B) \ \forall A \subset B \subset X$;

- Continuity: if $\{A_i\}_{1 \leq i}$ satisfies $A_i \subset X, A_{i+1} \subset A_i, \forall i$, then

$$\lim_{x \to \infty} g(A_i) = g\left(\bigcup_{i=1}^{\infty} A_i\right).$$

Fuzzy measures represent the importance of the set of properties X. They are defined not only in every property separately but also for all the aggregations among the properties of the set X.

Definition 2. *[25] A fuzzy measure is a Sugeno measure (or a $\lambda-$fuzzy measure) if satisfying the following additional conditions:*

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B).$$

For all $A, B \subset X, A \cap B = \emptyset$, and when $X = \{x_1, x_2, \ldots, x_n\}$ is a finite set, $X \neq \emptyset$, $\lambda > -1, \lambda \neq 0$, is determined by solving the following equation:

$$(\lambda + 1) = \prod_{i=1}^{n} \left(1 + \lambda g^{(i)}\right) \tag{1}$$

where, $g^{(i)} \stackrel{def}{=} g(\{x_i\})$. Thus, the fuzzy measure of a subset $A \subset X$ can be inferred from the following formula:

$$g(\{A\}) = \frac{-1 + \prod_{i=1}^{k}(1 + \lambda g^{(i)})}{\lambda} \tag{2}$$

where, $A = \{x_1, x_2, \ldots, x_k\}$.

Definition 3. *[16] Let X be a finite set, $X \neq \emptyset$. The Möbius transform of a given fuzzy measure g is a set function $\mathcal{M}^g : 2^X \to [0,1]$ defined by*

$$A \subset X \mapsto \mathcal{M}^g(A) = \sum_{B \subset A} (-1)^{|A \backslash B|} g(B). \tag{3}$$

The Möbius transformation is invertible. The fuzzy measure restored by its inverse transformation, called *Zeta transform*

$$g(A) = \sum_{B \subset A} \mathcal{M}^g(B), \forall A \subset X. \tag{4}$$

The boundary and monotonicity conditions of of a fuzzy measure are explained, respectively, as follows:

- $\mathcal{M}^g(\emptyset) = 0$ and $\sum_{A \subset X} \mathcal{M}^g(A) = 1$;

- $\sum_{B \subset A | i \in B} \mathcal{M}^g(B) \geq 0, \forall A \subset X, i \in A$.

Definition 4. *[23] A fuzzy measure g is called $\mathcal{K}$-additive if*

- $\forall A \subset X, \mathcal{M}(A) = 0$ if $card(A) > \mathcal{K}$;

- $\exists A \subset X$ such that $card(A) = \mathcal{K}$ and $\mathcal{M}^g(A) \neq 0$.

A $\mathcal{K}$-additive fuzzy measure is defined by only $\binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{k}$ coeficients instead of $2^n - 1$ of a general fuzzy measure g, where $n = card(X)$.

2.2.   Choquet integral

Choquet integral [24] is the most common form of fuzzy integral which is an integral of a real function with respect to a fuzzy measure [9].

Definition 5. *[25] Let $f : X \to [0,1]$ be a function where $f(x)$ denotes the confidence value of x. The Choquet integral of f with respect to the fuzzy measure g can be defined as:*

$$C_g(f) \overset{def}{=} \int_x f(x) \circ g \overset{def}{=} \int_0^1 g(A_\alpha) \, d\alpha \tag{5}$$

where, $\forall \alpha \in [0,1], A_\alpha \overset{def}{=} \{x \in X \,|\, f(x) \geq \alpha\}$. In case X is a discrete set, the discrete Choquet integral with respect to a fuzzy measure g is defined as

$$C_g(f) = \sum_{i=1}^n \left[f(x_{t_i}) - f(x_{t_{i-1}}) \right] g(A_{t_i}) \tag{6}$$

such that $0 \leq f(x_{t_1}) \leq f(x_{t_2}) \leq ... \leq f(x_{t_n}), f(x_{t_0}) = 0, A_{t_i} \overset{def}{=} \{x_{t_i}, ..., x_{t_n}\}$.

With the help of the Möbius transformation, Choquet integral can be expressed as

$$C_g(f) = \sum_{A \subset X} \mathcal{M}^g(A) \min_{i \in A} f(x_i). \tag{7}$$

3.   THE PROPOSED METHOD

In this section, the integral Choquet is used as aggregation operator to produce the overall similarity measure from individual similarity measures. The first, the form property of Choquet integral is clarified by the following statement:

Proposition 1.

(i) If g_λ is a $\lambda-fuzzy$ measure, then g_λ is monotonicity, $\forall A \subset B \subset X \Rightarrow g_\lambda(A) \leq g_\lambda(B)$,

and $A \neq \emptyset, A \subset X, A$ is finite $\Rightarrow g_\lambda(A) = \dfrac{-1 + \prod\limits_{a \in A} (1 + \lambda g_\lambda(\{a\}))}{\lambda}$.

(ii) If $f_1, f_2 : X \to [0,1], f_1(x) \leq f_2(x), \forall x \in X$, g is the fuzzy measure on X then $C_g(f_1) \leq C_g(f_2)$.
Special, $\inf(f) \leq C_g(f) \leq \sup(f), \forall f : X \to [0,1]$. This property demonstrates that when X is a finite set, the C_g is an aggregation operator.

(iii) $X = \{ x_i \,|\, i = \overline{1,n} \}$, g is the fuzzy measure on X, $f : X \to [0,1]$ is a function and $i_1, i_2, \ldots, i_n$ is a permutation of $1, 2.., n$ such that: $0 = f(x_{i_0}) \leq f(x_{i_1}) \leq f(x_{i_2}) \leq ... \leq f(x_{i_n}), \forall k = \overline{1,n}, A_{i_k} \overset{def}{=} \{ x_{i_k}, ..., x_{i_n} \}, A_{i_{n+1}} \overset{def}{=} \emptyset$, then

$$\sum_{k=1}^{n} \left(f(x_{i_k}) - f(x_{i_{k-1}}) \right) g(A_{i_k}) = \sum_{k=1}^{n} f(x_{i_k}) \left(g(A_{i_k}) - g(A_{i_{k+1}}) \right).$$

Proof:

(i) $A \subset B \Rightarrow B = A \cup B \backslash A \Rightarrow g_\lambda(B) = g_\lambda(A) + g_\lambda(B \backslash A) + g_\lambda(B) g_\lambda(B \backslash A)$

$= g_\lambda(A) + g_\lambda(B \backslash A) \{ 1 + \lambda g_\lambda(A) \} \geq g_\lambda(A)$ by $0 \leq g_\lambda(B \backslash A), 0 \leq g_\lambda(A) \leq 1, \lambda > -1$.

Rewrite the Sugeno equation:
$\forall A, B \subset X, A \cap B = \emptyset, g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B)$
$\Rightarrow (1 + \lambda g_\lambda(A \cup B)) = (1 + \lambda g_\lambda(A))(1 + \lambda g_\lambda(B))$
$\Rightarrow (1 + \lambda g_\lambda(\{a_1, a_2, .., a_i\})) = (1 + \lambda g_\lambda(\{a_1, a_2, .., a_{i-1}\}))(1 + \lambda g_\lambda(\{a_i\}))$,

$\forall i = k, k-1, ..., 2$

$\Rightarrow 1 + \lambda g_\lambda(\{a_1, a_2, .., a_k\}) = \prod\limits_{i=1}^{k} (1 + \lambda g_\lambda(\{a_i\}))$

$\Rightarrow g_\lambda(\{a_1, a_2, .., a_k\}) = \dfrac{-1 + \prod\limits_{i=1}^{k} (1 + \lambda g_\lambda(\{a_i\}))}{\lambda}$

(ii) $\forall \alpha \in [0,1], \{ x \in X \,|\, f_1(x) \geq \alpha \} \subset \{ x \in X \,|\, f_2(x) \geq \alpha \}$ do $f_1(x) \leq f_2(x) \forall x \in X$

$$\Rightarrow g(\{ x \in X \,|\, f_1(x) \geq \alpha \}) \leq g(\{ x \in X \,|\, f_2(x) \geq \alpha \}) \forall \alpha \in [0,1]$$

$$\Rightarrow C_g(f_1) = \int_0^1 g(\{ x \in X \,|\, f_1(x) \geq \alpha \}) d\alpha \leq \int_0^1 g(\{ x \in X \,|\, f_2(x) \geq \alpha \}) d\alpha = C_g(f_2)$$

$$\underline{f}: X \to [0,1], \forall x \in X, \underline{f}(x) = \inf(f), \overline{f}: X \to [0,1], \forall x \in X, \overline{f}(x) = \sup(f)$$

$$\inf(f) = C_g(\underline{f}) \le C_g(f) \le C_g(\overline{f}) = \sup(f)$$

(iii) $\forall k = \overline{1,n}, a_k \overset{def}{=} f(x_{i_k}) - f(x_{i_{k-1}}), s_0 = 0, s_k \overset{def}{=} \sum_{l=1}^{k} a_l \Rightarrow s_k = f(x_{i_k})$

$$\sum_{k=1}^{n} \left(f(x_{i_k}) - f(x_{i_{k-1}}) \right) g(A_{i_k}) = \sum_{k=1}^{n} a_k g(A_{i_k}) = \sum_{k=1}^{n} s_k g(A_{i_k})$$

$$= f(x_{i_n})g(A_{i_{n+1}}) + \sum_{k=1}^{n} s_k \left\{ g(A_{i_k}) - g(A_{i_{k+1}}) \right\} (Abel's\ formular\ )$$

$$= \sum_{k=1}^{n} f(x_{i_k}) \left\{ g(A_{i_k}) - g(A_{i_{k+1}}) \right\}.$$

When g is the additive fuzzy measure,

$$C_g(f) = \sum_{k=1}^{n} f(x_{i_k}) \left\{ g(A_{i_{k-1}}) - g(A_{i_k}) \right\} = \sum_{k=1}^{n} f(x_{i_k}) g(\{x_{i_k}\}) = \sum_{i=1}^{n} f(x_i) g(\{x_i\}).$$

∎

Remaks:

- *Formula $g_\lambda(A)$ in (i) demonstrates that all pairs of features interact the same way, is a polynomial of λ in which coefficients are importance of individual feature.*
  $A = \{a_1, a_2, .., a_k\} \Rightarrow$
  $g_\lambda(A) = \sum_{i=1}^{k} g_\lambda(\{a_i\}) + \lambda \sum_{i_1=1}^{k-1} \sum_{i_2=i_1+1}^{k} g_\lambda(\{a_{i_1}\}) g_\lambda(\{a_{i_2}\}) + ... + \lambda^{k-1} \prod_{i=1}^{k} g_\lambda(\{a_i\}).$

- Right side of the equation (iii) means that, if new information that x_{i_k} contribution is less, the measure should be chosen so that $g(A_{i_k}) - g(A_{i_{k+1}}) \approx 0$.

3.1.   Modeling the overall similarity using the discrete Choquet Integral

Let $F = \{f_t\}_{t=1}^{T}$ be a set of n feature spaces which represent an image used by CBIR system. Here, f_t is a multi-dimension vector of feature space which can be modelled by several representations, e.g. Color Histogram, Grid Color Moment, Local Binary Pattern, Gabor Wavelets Texture and Edge Histogram. Let F^q and F^j be two feature vectors of two images q, j, respectively.

Definition 6. *The similarity measure with respect to multi-feature between two images q and j, is denoted as $Sim^T(q,j)$ and is defined as follows:*

$$Sim^T(q,j) = \Phi \left(\left\{ Sim_t \left(f_t^q, f_t^j \right) | t = 1...T \right\} \right) \tag{8}$$

where, T is number of feature spaces, $Sim_t\left(f_t^q, f_t^j\right)$ is similarity measure between query image and image j according to t^{th} feature space, Φ is a paradigm for aggregating the similarity measure of individual feature spaces into the overall similarity measure. This definition is general for using many types of different feature spaces. The most popular aggregation paradigm, which is used in CBIR systems, is the weighted sum. This paradigm is defined as follows:

$$\Phi\left(\left\{Sim_t\left(f_t^q, f_t^j\right) | t = 1...T\right\}\right) = \sum_{t=1}^{T} \lambda_t\, Sim_t\left(f_t^q, f_t^j\right) \tag{9}$$

where, $\sum_{t=1}^{T} \lambda_t = 1$, and $\lambda_t \geqslant 0, \forall t = \overline{1, T}$.

In practice, when users observe the images, they can recognize the similar between two images in either an individual feature or in some of aggregating features. In addition, a similar of individual feature is only important for the user if similarity with respect to some other feature appears, it means that, there is a mutual interaction in each other features. Thus, above model is not flexibly enough for such query paradigms. For example, assuming that there has three images I_1, I_2, I_3 and their overall similarities are evaluated based on three feature namely *color*, *texture* and *shape*, and the weight is assigned on each feature to be the same. Using the aggregation paradigm in Eq(9), the overall similarity measure of images is shown in Table 1. In this results, image I_3 has the lowest rank. However, if the user wants to search for a well-balanced both *shape* and *texture* similarity, image I_3 should be considered as the first rank. This will be difficult to achieve by additive operators, although the λ changes with any values.

Table 1. The overall similarity using weighted sum operator
(with weights $\lambda(color) = \lambda(texture) = \lambda(shape) = 1/3$)

Image	Color	Texture	Shape	Overall Similarity
I_1	0.72	0.68	0.44	0.613
I_2	0.60	0.52	0.72	0.613
I_3	0.56	0.60	0.60	0.587

In this case, the weights should be assigned to both the individual feature and their subsets. For above example, the weight assigned to the aggregation of two featuresbreak $\{texture, shape\}$ should be bigger than or equal to the sum of their individual weights, i.e.,

$$\lambda\left\{texture, shape\right\} \geq \lambda\left\{texture\right\} + \lambda\left\{shape\right\}.$$

Additionally, the weights assigned to the rest combination should be smaller or equal to the sum of their individual weight because they do not receive much interest of users, i.e., $\lambda\left\{color, texture\right\} \leq \lambda\left\{color\right\} + \lambda\left\{texture\right\}$, similarly, yielding $\lambda\left\{color, shape\right\} \leq \lambda\left\{color\right\} + \lambda\left\{shape\right\}$. This reasoning can be easy to model by fuzzy measures [15]. Hence, the Choquet integral will be effective solution for the query paradigms.

In this work, Choquet integral with fuzzy measure is used to define aggregation paradigm Φ. The similarity of features are considered as source information, and the relevance weight

of features is considered as fuzzy density function. These are then used in Choquet integral model to compute the overall similarity measure. Thus Eq(8) can be rewritten as follows:

$$
\begin{aligned}
\mathrm{Sim}^T(q,j) &= Choquet_\mu(q,j) \\
&\stackrel{\mathrm{def}}{=} \sum_{i=1}^{T} \left(Sim_{t_{(i)}}\left(f_t^q, f_t^j\right) - Sim_{t_{(i-1)}}\left(f_t^q, f_t^j\right) \right) \times \mu\left(\{f_i, ..., f_T\}\right)
\end{aligned}
\tag{10}
$$

where $\{t_i\}_{1 \le i \le T}$, it means that, the indexes of features set are permuted in such a way that: $Sim_{t_1}\left(f_t^q, f_t^j\right) \le ... \le Sim_{t_{(T)}}\left(f_t^q, f_t^j\right)$, and convention $Sim_{t_0}\left(f_t^q, f_t^j\right) = 0$. The notation $\mu(.)$ presents the relevance weights of features which will also serve as the with fuzzy measures under the constraints as:

$$
\begin{cases}
\{\mu(A)\}_{A \in 2^{\{1,2,..,T\}}}, \forall A \in 2^{\{1,2,...,T\}}, 0 \le \mu(A) \le 1 \\
\mu(\emptyset) = 0, \mu\left(\{1,2,...T\}\right) = 1 \\
\forall A \in 2^{\{1,2,...,T\}}, A \ne \emptyset, \displaystyle\max_{B \subset A, |B| = A - 1} \{\mu(B)\} \le \mu(A).
\end{cases}
\tag{11}
$$

Using $\mathcal{K}$-additive fuzzy measure in Möbius representation, formula Eq(10) is able to be rewritten as follows:

$$
Sim^T(q,j) \stackrel{\mathrm{def}}{=} \sum_{A \subseteq \{1,...,T\}} \mathcal{M}_A^\mu \min_{t \in A}\left(Sim_t\left(f_t^q, f_t^j\right) \right)
\tag{12}
$$

where $\mathcal{M}^\mu(A)$ satisfying constrains:

$$
\begin{cases}
\sum_{B \subseteq A, |B| \le K} \mathcal{M}_B^\mu \ge 0 \\
\forall A \subseteq \{1,2,..,T\}, |A| > 1, \forall t \in A, \mathcal{M}_{\{t\}}^\mu \ge 0 \\
\sum_{B \subseteq \{1,2,...,T\}, |B| < \mathcal{K}} \mathcal{M}_B^\mu = 1.
\end{cases}
\tag{13}
$$

Considering above example, suppose the individual importance of each feature is assigned by fixed values where $\mu\{color\} = \mu\{texture\} = \mu\{shape\} = 0.3$, and the importance on the aggregation of features is assigned by:

$$
\mu\{textue, shape\} = 0.9 \ge \mu\{texture\} + \mu\{shape\},
$$

$$
\mu\{color, texture\} = 0.4 \le \mu\{color\} + \mu\{texture\},
$$

$$
\mu\{color, shape\} = 0.4 \le \mu\{color\} + \mu\{shape\}.
$$

According to Definition 1, $\mu\{\emptyset\} = 0$, $\mu\{color, texture, shape\} = 1$. The fuzzy measure values and similarity measures in Table 1 are substituted into Eq(10) to compute the overall similarity. The result is summarized as shown in Table 2. Based on this result, we see, image I_3 is identified as the best similar image which fit user's expectation.

Table 2. The overall similarity using integral Choquet

Image	Overall similarity measure	Ranking
I_1	*Step1*: Sorting the feature's similarity measure in ascending order: $\text{Sim}_s < \text{Sim}_t < \text{Sim}_c$. *Step2*: Calculating the overrall similarity measure according using Eq(10) C_g(color, texture, shape)= Sim_s*g{color,texture,shape}+ (Sim_t-Sim_s)*g{texture,color}+(Sim_c-Sim_t)*g{color} =0.44*1+(0.68-0.44)*0.4+(0.72-0.68)*0.2 =0.548	3
I_2	0.588	2
I_3	0.596	1

3.2. Learning features's relevance weight

To use the Choquet integral for aggregating similarity, it is necessary to identify the weight of all subsets of feature space. However, it is difficult for estimating 2^n values of fuzzy measure, especially when the number of features, n is sufficiently large [27]. Many methods have been proposed to identify fuzzy measure values, the most of them are stated under of the form of an optimization problem: see [7] for detail. However, finding a solution via these methods in the case of a very large number of parameters and constraints still remains as an obstacle. In general, the usability of each identification method depending on several factors: types of inputs, number of inputs and number of fuzzy measure values that need to be identified. In this work, the learning relevance weight of features is formulated in the context of linear programming problem, which can be solved quickly and reliably.

Let S^+, S^- be two sets of image related and unrelated, which labeled by user via relevance feedback. In this work, it is needed to estimate relevance weights μ of candidate subsets of features so that the value of the overall similarity match the desired outputs y_k as close as possible, i.e., that the images in S^+ are ranked in closest to query image, whereas the images in S^- are ranked in farthest to query image. For this purpose, the learning feature's relevance weight can be formulated as the least absolute deviation problem:

$$\text{minimize} \sum_{k=1}^{K} |Choquet_\mu(q,j) - y_k| \qquad (14)$$

subject to various constraints on μ.

Let $y_k = \left\{ y_k^+, y_k^- \right\}$ be the output values which are defined by

$$y_k^+ = \max_{1 \le t \le T} Sim_t(f_t^q, f_t^j), I_k \in S^+,$$
$$y_k^- = \min_{1 \le t \le T} Sim_t(f_t^q, f_t^j), I_k \in S^-. \qquad (15)$$

The problem (14) can be restated as follows:

$$\text{minimize} \sum_{p \in S^+} \max \left(y_p^+ - Choquet_\mu \left(p, j \right), 0 \right)$$
$$+ \sum_{p \in S^-} \max \left(Choquet_\mu \left(p, j \right) - y_p^-, 0 \right) \tag{16}$$

subject to various constraints (11) on μ. However, these conditions do not reduce the complexity of the least absolute deviation problems. It is possible to reduce the complexity of the problem when working in Möbius representation [6].

By using $\mathcal{K}$-additive fuzzy measure in Möbius representation, the problem (16) is transformed into a simplified optimization problem:

$$\text{minimize} \sum_{p \in S^+} \max \left(y_p^+ - \sum_{A||A|\leq \mathcal{K}} \mathcal{M}_A^\mu h_A \left(p \right), 0 \right)$$
$$+ \sum_{p \in S^-} \max \left(\sum_{A||A|\leq \mathcal{K}} \mathcal{M}_A^\mu h_A \left(p \right) - y_p^-, 0 \right) \tag{17}$$

subject to various constraints:

$$\begin{cases} \sum_{B \subseteq A, |B| \leq K} \mathcal{M}_B^\mu \geq 0 \\ \forall A \subseteq \{1, 2, .., \mathrm{T}\}, |A| > 1, \forall t \in A, \mathcal{M}_{\{t\}}^\mu \geq 0 \\ \sum_{B \subseteq \{1,2,...,\mathrm{T}\}, |B| < \mathcal{K}} \mathcal{M}_B^\mu = 1 \end{cases} \tag{18}$$

where

$$h_A \left(p \right) = \min_{t \in A} \left(Sim_t \left(f_t^q, f_t^p \right) \right),$$

Set

$$e_p^+ = y_p^+ - \sum_{A||A|\leq \mathcal{K}} h_A(p)\mathcal{M}_A^\mu, p \in S^+,$$

$$e_p^- = \sum_{A||A|\leq \mathcal{K}} h_A(p)\mathcal{M}_A^\mu, -y_p^-, p \in S^-.$$

By using auxiliary variables e_p^+, e_p^-, the problem (17) is transformed into:

$$\text{minimize} \sum_{p=1}^{K} e_p^+ + e_p^- \tag{19}$$

subject to constraints:

$$\begin{cases} e_p^+ - \sum_{A||A|\leq \mathcal{K}} h_A(p)\mathcal{M}_A^\mu \geq -y_p^+, \forall p \in S^+, \\ e_p^- + \sum_{A||A|\leq \mathcal{K}} h_A(p)\mathcal{M}_A^\mu \geq y_p^-, \forall p \in S^-, \\ \sum_{B \subseteq A, |B| \leq \mathcal{K}} \mathcal{M}_B^\mu \geq 0, \\ \forall A \subseteq \{1, 2, .., \mathrm{T}\}, |A| > 1, \forall t \in A, \mathcal{M}_{\{t\}}^\mu \geq 0, \\ \sum_{B \subseteq \{1,2,...,\mathrm{T}\}, |B| < \mathcal{K}} \mathcal{M}_B^\mu = 1, \\ e_p^+ \geq 0, e_p^- \geq 0. \end{cases} \tag{20}$$

This is a linear programming problem which is very convenient to solve by using several standard methods as Simplex method. The procedure for calculating overall similarity measure is described in algorithm 1, and the proposed method for content-based image retrieval is described in algorithm 2.

Algorithm 1 Aggregation similariy measure based on Choquet integral

---

Input: Set of m fuzzy measure in Möbius representation $Mob = \left\{ \mathcal{M}_A^\mu, \forall A \subseteq \{1, 2, .., \mathrm{T}\} \right\}$,

  set of T similarity measures $X = \left\{ Sim_t \left(f_t^q, f_t^j \right) \right\} | t = 1...T$.

Output: aggregation similarity measure S.

  1: *Initialisation*: $S = 0$;
  2: **for** $j = 1$ to m **do**
  3:   **for** $i = 1$ to T **do**
  4:     r=0;
  5:     **if** bit i of j ($from\ right\ to\ left$) is 1 **then**
  6:       $r = min(r, X[i])$;
  7:     **end if**
  8:   **end for**
  9:   **if** $r > 1$ **then**
  10:     $r = 0$;
  11:   **end if**
  12:   $S = S + Mob[j] * r$;
  13: **end for**
  14: **return** S;

---

4. EXPERIMENTS

4.1. Image database

In this section, the efficiency of proposed method is we evaluated on two sets of data as ImgDB1 and ImgDB2 is evaluated. The ImgDB1 is set of Caltech101 data [12] including 9144 images classified into 101 classes, there are from 40 to 800 images on each one. The ImgDB2 is subset of set of Corel data [29] including 15000 images classified into 150 classes with different semantic topics. There are 100 images on each one. They are benchmark data set that are often used to evaluate the efficiency in CBIR systems [11, 29].

4.2. Feature extraction

In CBIR system in general, each image is often presented by main three features: Color, Texture and Shape [22, 11]. In the experiment, five descriptions for these features in [36] are used, that are: a 9-dimension-vector for Grid Color Moment (GCM), a 59-dimension-vector for Local Binary Pattern (LBP), a 120-dimension-vector for Gabor Wavelets Texture (GWT), a 37-dimension-vector for Edge Histogram (EH) and a 512-dimension vector for GIST feature. All of feature descriptions have been normalized according to method [1] so that their each element has value in range [0, 1]. The similarity measure between pairs of

---

Algorithm 2 CBIR-Choquet (Proposed method for content-based image retrieval)

Input: Query image $x^q = \{f_t^q\}_{1 \le t \le T}$, Feature database DB$= \{x^j\}_{1 \le j \le N}$, $x^j = \{f_t^j\}_{1 \le t \le T}$,

  where, $f_t^q, f_t^j \in [0,1]$.

Output: Result: Image list are arranged in the similarity measure to query image

 1: *Initialisation*: $g(Sim_t) = \frac{1}{T} \forall t = \overline{1,T}; S^+ = \emptyset; S^- = \emptyset;$
   % *Ranking without relevance feedback*

 2: **for** $j = 1$ to N **do**

 3:   Compute the similarity between the query image and every image in DB according to feature spaces. $\left\{ Sim_t \left(f_t^q, f_t^j \right) \right\} |t = 1...T$

 4:   Compute overall similarity between the query image and every image in DB $Sim^T DB(j) = Sim^T(q,j);$ using Eq(9)

 5: **end for**

 6: Sort $Sim^T DB(j), 1 \le j \le N;$ in descending order

 7: Normalize $\left\{ Sim_t \left(f_t^q, f_t^j \right) \right\} |t = 1...T$, s.t $Sim_t \left(f_t^q, f_t^j \right) \in [0,1].$
   % *In each kth round of relevance feedback*

 8: **while** User not satisfied **do**

 9:   Identify sets of S_k^+, S_k^- from user's feedback:
     $S^+ = S^+ \cup S_k^+; S^- = S^- \cup S_k^-;$

10:   Compute the $y_k^+, y_k^-;$ using Eq(15)

11:   Compute the $\mathcal{M}_A^\mu;$ by solving the problem (19).

12:   **for** $j = 1$ to N **do**

13:     Compute the overall similarity
       $Sim^T DB(j) = Choquet(\left\{ Sim_t \left(f_t^q, f_t^j \right) \right\} |t = 1...T, \mathcal{M}_A^\mu);$ by algorithm 1

14:   **end for**

15:   *Result*= Sort $Sim^T DB(j), 1 \le j \le N;$ in descending order

16: **end while**

17: **return** *Result*;

---

image of each description is calculated by using Euclidean distance. The descriptions used in experiment and their properties are shown in Table 3.

4.3.   Evaluating efficiency

A serial of experiments is implemented to show the feasibility and utility of the proposed method. To illustrate the specific situation of online users, twenty images are selected randomly from each class of each image database and considered as query images. Therefore, it will have 2020 and 3000 query sessions on each database, respectively. At initial step of each query session, images in database are sorted by Euclid distance to query image, and then RF is automatically performed by a computer. At each feed-back round, the top 20 retrieved results are serially examined from the top, and all the unlabeled images within the scope were automatically labeled using the ground truth in order to simulate the users feedback. The images which belong to the same class are considered relevance and others are considered irrelevance. All of labeled images in feedback loops will be used to update

Table 3. Descriptions and their properties used in the experiment

Descriptions	Type of feature	Size of feature vectors
GCM	Color	81
LBP	Texture	59
GWT	Texture	120
EDH	Shape	37
GIST	Shape	512

fuzzy weights that are used to combine similarity measure.

The query result without feedback on set of data ImgDB1 is shown in Fig.1, distance measure values are shown in Table 4, respectively. The image at the left top position is query one, the images which have red border are the same semantic class with query image, others are different from semantic class. It is very easy to see that amount of images having relevance to query image is very limited; a lot of images have a very closed distance to query image but very different in semantic and vice versa. However, after the fourth feedback iteration, number of relevance images which are retrieved by proposed method increases significantly as Fig.2.

Figure 1. The retrieval results without relevance feedback for query image 377022.*jpg*.

The Average Precision measure (AP)[18] has used to evaluate efficiency of the system after each feedback iteration. At each feedback iteration, AP is defined as the average of precision values obtained after each relevant image is retrieved. The precision value is the ratio between the number of relevant images found by the system and the number of pictures currently retrieved at the current iteration. The mean average precision (MAP) is the mean of the average precision over all queries:

$$MAP = \frac{1}{Q} \sum_{q=1}^{Q} \frac{1}{N_R} \sum_{i=1}^{N_R} p(i) \tag{21}$$

Figure 2. The retrieval results after the fourth feedback iteration for query image 377022.*jpg*.

Table 4. Distance measure of the top 20 retrieved results without relevance feedback for query image 377022.*jpg*.

Image	GCM	LBP	GWT	EDH	GIST	Dist
377022	**0**	**0**	**0**	**0**	**0**	**0**
113006	0.241	0.081	0.149	0.214	0.357	1.042
377024	**0.225**	**0.112**	**0.174**	**0.216**	**0.373**	**1.099**
208035	0.166	0.078	0.198	0.235	0.425	1.101
377029	**0.236**	**0.113**	**0.170**	**0.227**	**0.363**	**1.110**
134011	0.233	0.094	0.208	0.233	0.349	1.117
377075	**0.213**	**0.176**	**0.183**	**0.225**	**0.326**	**1.124**
104079	0.193	0.110	0.151	0.318	0.355	1.128
377078	**0.256**	**0.109**	**0.142**	**0.237**	**0.387**	**1.130**
499011	0.228	0.108	0.219	0.188	0.393	1.136
136016	0.295	0.122	0.143	0.222	0.358	1.140
482032	0.285	0.092	0.123	0.264	0.383	1.147
113032	0.286	0.143	0.147	0.258	0.319	1.152
150022	0.301	0.115	0.177	0.249	0.317	1.160
77088	0.242	0.128	0.226	0.267	0.300	1.163
377035	**0.222**	**0.175**	**0.236**	**0.215**	**0.321**	**1.171**
283069	0.174	0.149	0.187	0.264	0.398	1.172
283056	0.208	0.119	0.233	0.223	0.390	1.172
208096	0.259	0.125	0.216	0.226	0.349	1.174
326018	0.213	0.102	0.190	0.239	0.434	1.177

where, Q is total of query image, N_R is the total number of relevant images for the query q.

The effectiveness of the proposed method is evaluated in two aspects: The effectiveness of weight learning model, and the aggregation model of similarity measures. With the first aspect, the proposed method is compared with the method in [13] which uses the Choquet

integral with Sugeno measure as the distance measure of features. In this method, the authors have used the Euclidean distance of each component in 48-dimensional feature vector representing the image's texture features as sources of information in the CI. The coefficient λ that is used to compute weight of interactive features is computed by estimating the real roots of a polynomial of degree 48. However, in the context of multi-features, applying this method directly for each feature component is not suitable due to the need of estimating the real root of the polynomial of degree $n(n = 809$ in our case) when calculating the coefficient λ. In this experiment, the information source used in Choquet integral is the similarity measure of five feature tuples, and the weight of the interactive features is evaluated according to the respective proposals. Determining of the coefficient λ is easily performed by finding the real root of a polynomial of degree 5.
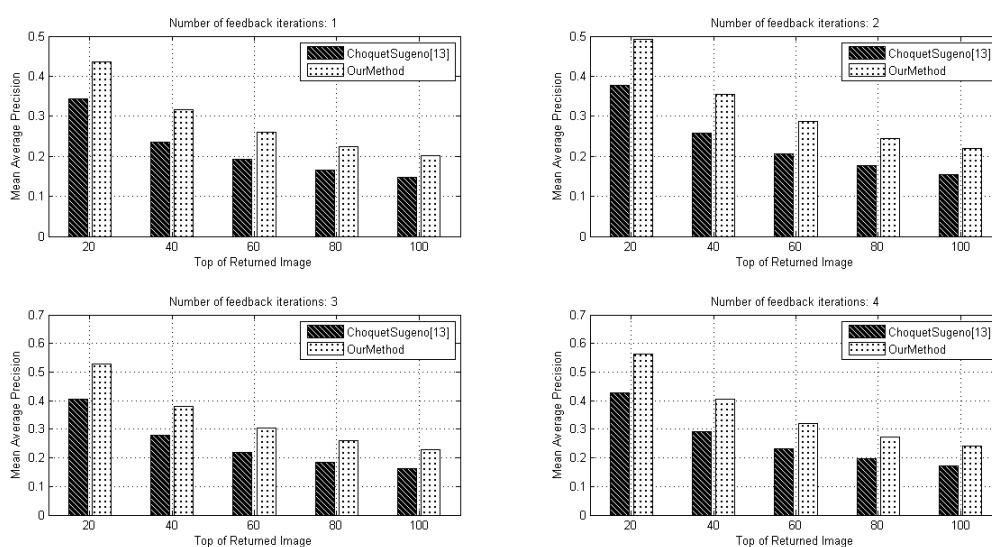

Figure 3. The effectiveness of learning model of fuzzy measure on ImgDB1 dataset

The results in Fig.3 and Fig.4 show that the proposed method is superior to the method using Sugeno measure. The reason is that, in the proposed method, the weight of the interactive features is more closely reflected user's target because this weight is learned during the relevant feedback process. Meanwhile, the method proposed in [13] only learns the important weights of the corresponding individual feature tuple. These weights are then used to estimate the coefficient λ that is used to compute the weight of the interactive features, meaning that this method only considers the interaction of all pairs of feature tuples in the same way, which cannot fully reflect the user's target.

To evaluate the performance of the proposed method according to the second aspect, an experiment is carried out as follows: Firstly, the method [33] is used to compute similarity measure for individual feature tuple. Then, these similarity measures are aggregated according to three ways: using weighted sum operator, our proposed method, and the method proposed in [4] which are named EMR+WS, EMR+Proposed and IMR+[4], respectively. The weights of the feature tuples are estimated according to the learning algorithm in the proposed method. The comparison results are shown in Fig.5. In most cases, the results of
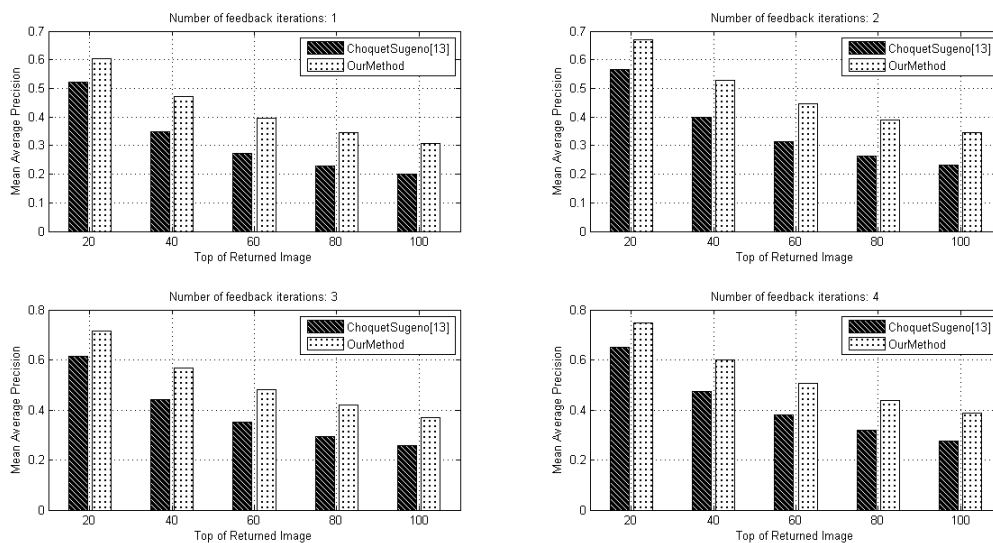
Figure 4. The effectiveness of learning model of fuzzy measure on ImgDB2 dataset
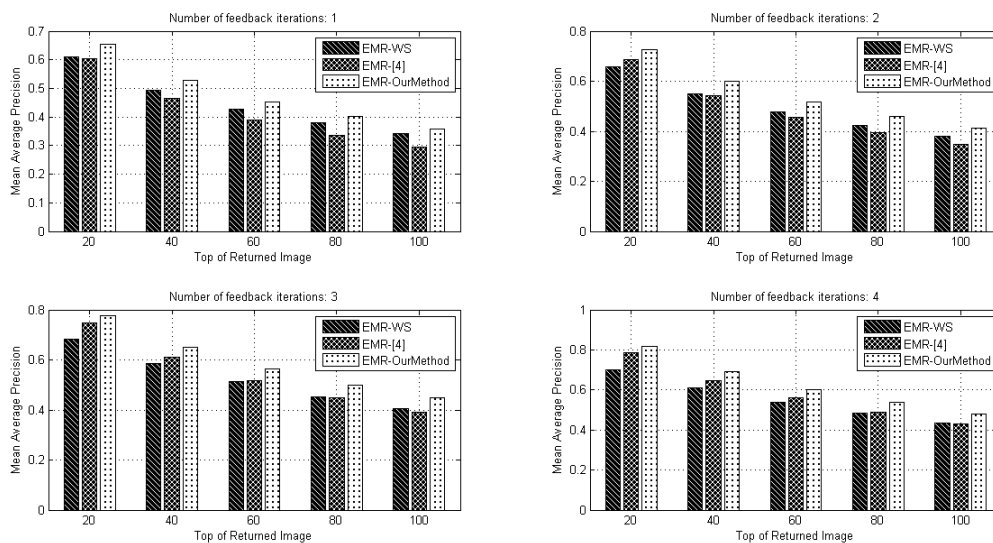

Figure 5. The effectiveness of aggregation paradigm of similarity measures on ImgDB2 dataset

the proposed method are better than those of other methods. As analyzed above, the aggregation model based on Choquet integral considers all the interactions between subsets of the feature tuples, so the results of this model reflect more accurately the target of users, especially, in the case of multiple semantic concepts, which is present in the image as in ImgDB2. The computation of weights of interactive features is expensive due to the need of considering all the interactions. However, the proposed method applies aggregation paradigm of

similarity measures for T feature tuples ($T = 5$ in this case) instead of computing individual feature components, so this difficulty can be easily overcome.

5. CONCLUSION

This paper proposes the method to aggregate multi-similarity measures based on relevance feedback in CBIR. In this approach, the overall similarity measure between the two images is calculated by non-additive aggregation from different individual similarity measure based on Choquet integral. The important weights are used in Choquet integral, which are defined not only for each individual similarity, but also for all the aggregations between them. The weights are estimated by the Simplex algorithm for optimization linear objective function based on user's feedback information, therefore, the overall similarity measure is more suitable than the target user's query. Experimental results on two benchmark datasets Oliva and Corel have shown the effectiveness of the proposed approach compared to two methods [13] and [4]. Despite promising results, identifying the most informative images for user to label, which is one of the critical issues in relevance feedback has not been addressed in the proposed method yet. The authors plan to incorporate active learning methods in the proposed to more effectively exploit the feedback information of users.

ACKNOWLEDGMENT

REFERENCES

[1] S. Aksoy and R. M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 563–582, 2001, image/Video Indexing and Retrieval.

[2] M. Arevalillo-Herráez, F. J. Ferri, and J. Domingo, "A naive relevance feedback model for content-based image retrieval using multiple similarity measures," *Pattern Recogn.*, vol. 43, no. 3, pp. 619–629, Mar. 2010.

[3] M. Arevalillo-Herrez, J. Domingo, and F. J. Ferri, "Combining similarity measures in content-based image retrieval," *Pattern Recognition Letters*, vol. 29, no. 16, pp. 2174–2181, 2008.

[4] M. Arevalillo-Herrez, F. J. Ferri, and J. Domingo, "A naive relevance feedback model for content-based image retrieval using multiple similarity measures," *Pattern Recognition*, vol. 43, no. 3, pp. 619–629, 2010.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008, similarity Matching in Computer Vision and Multimedia.

[6] G. Beliakov, "Fitting fuzzy measures by linear programming. programming library fmtools," in *Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on*, June 2008, pp. 862–867.

[7] W. Budiharto, A. R. Krishnan, M. M. Kasim, and E. M. N. E. A. Bakar, "A short survey on the usage of choquet integral and its associated fuzzy measure in multiple attribute analysis," *Procedia Computer Science*, vol. 59, pp. 427–434, 2015.

[8] O. M. A. B. Chawki Youness, El Asnaoui Khalid, "New method of content based image retrieval based on 2-d esprit method and the gabor filters," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 15, no. 2, pp. 313–320, August 2015.

[9] Y. Choi, D. Kim, and R. Krishnapuram, "Relevance feedback for content-based image retrieval using the choquet integral," in *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, vol. 2, 2000, pp. 1207–1210 vol.2.

[10] R. da S. Torres, A. X. Falcão, B. Zhang, W. Fan, E. A. Fox, M. A. Gonçalves, and P. Calado, "A new framework to combine descriptors for content-based image retrieval," in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, ser. CIKM '05.   New York, NY, USA: ACM, 2005, pp. 335–336.

[11] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1–60, May 2008.

[12] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Comput. Vis. Image Underst.*, vol. 106, no. 1, pp. 59–70, Apr. 2007.

[13] H. Frigui, "Interactive image retrieval using fuzzy sets," *Pattern Recognition Letters*, vol. 22, no. 9, pp. 1021–1031, 2001.

[14] N. T. Giang, N. Q. Tao, N. D. Dung, and N. T. The, "Skeleton based shape matching using reweighted random walks," in *The proceding of the IEEE on 9th International Conference on Information, Communications and Signal Processing (ICICS)*, December 2013, pp. 1–5.

[15] M. Grabisch, "The application of fuzzy integrals in multicriteria decision making," *European Journal of Operational Research*, vol. 89, no. 3, pp. 445–456, 1996.

[16] M. Grabisch, I. Kojadinovic, and P. Meyer, "A review of methods for capacity identification in choquet integral based multi-attribute utility theory: Applications of the kappalab r package," *European Journal of Operational Research*, vol. 186, no. 2, pp. 766–785, 2008.

[17] W. Hu, N. Xie, L. Li, X. Zeng, and S. Maybank, "A survey on visual content-based video indexing and retrieval," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 6, pp. 797–819, Nov 2011.

[18] M. J. Huiskes and M. S. Lew, "Performance evaluation of relevance feedback methods," in *Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval*, ser. CIVR '08.   New York, NY, USA: ACM, 2008, pp. 239–248.

[19] Q. Iqbal and J. K. Aggarwal, "Combining structure, color and texture for image retrieval: A performance evaluation," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 2, 2002, pp. 438–443 vol.2.

[20] J. D. Knowles and D. W. Corne, "Approximating the nondominated front using the pareto archived evolution strategy," *Evol. Comput.*, vol. 8, no. 2, pp. 149–172, Jun. 2000.

[21] C.-H. Lee and M.-F. Lin, "Ego-similarity measurement for relevance feedback," *Expert Systems with Applications*, vol. 37, no. 1, pp. 871–877, 2010.

[22] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recogn.*, vol. 40, no. 1, pp. 262–282, Jan. 2007.

[23] G. Michel, "K-order additive discrete fuzzy measures and their representation," *Fuzzy Sets Syst.*, vol. 92, no. 2, pp. 167–189, Dec. 1997.

[24] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measures and the choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets and Systems*, vol. 29, no. 2, pp. 201–227, 1989.

[25] Y. Narukawa and T. Murofushi, *Choquet integral and Sugeno integral as aggregation functions.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 27–39.

[26] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 644–655, Sep 1998.

[27] A. Silvia, G. Salvatore, L. Fabio, and M. Benedetto, "Assessing non-additive utility for multi-criteria decision aid," *European Journal of Operational Research*, vol. 158, no. 3, pp. 734–744, 2004.

[28] M. S. T. Murofushi, "An interpretation of fuzzy measure and the choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets and Systems*, vol. 29, pp. 201–227, 1989.

[29] D. Tao, X. Tang, X. Li, and Y. Rui, "Direct kernel biased discriminant analysis: a new content-based image retrieval relevance feedback algorithm," *IEEE Transactions on Multimedia*, vol. 8, no. 4, pp. 716–727, Aug 2006.

[30] B. Thomee and M. Lew, "Interactive search in image retrieval: a survey," *International Journal of Multimedia Information Retrieval*, vol. 1, no. 2, pp. 71–86, 2012.

[31] J. Wu and J. M. Rehg, "Centrist: A visual descriptor for scene categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1489–1501, Aug 2011.

[32] L. Wu and S. C. H. Hoi, "Enhancing bag-of-words models with semantics-preserving metric learning," *IEEE MultiMedia*, vol. 18, no. 1, pp. 24–37, Jan 2011.

[33] B. Xu, J. Bu, C. Chen, C. Wang, D. Cai, and X. He, "Emr: A scalable graph-based ranking model for content-based image retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 1, pp. 102–114, Jan 2015.

[34] M. O. Y. Rui, T. S. Huang and S. Mehrotra, "Relevance feedback: A powerful tool for interactive content-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 644– 655, 1998.

[35] Q. Zhang and E. Izquierdo, "Optimizing metrics combining low-level visual descriptors for image annotation and retrieval," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 2, May 2006.

[36] J. Zhu, S. C. Hoi, M. R. Lyu, and S. Yan, "Near-duplicate keyframe retrieval by nonrigid image matching," in *Proceedings of the 16th ACM International Conference on Multimedia*, ser. MM '08. New York, NY, USA: ACM, 2008, pp. 41–50.

[37] Y. K. J. K. Zukuan WEI, Hongyeon KIM, "An efficient content based image retrieval scheme," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 11, no. 11, pp. 6986–6991, November 2013.