

# Earth Mover's Distance Region-based Image Similarity Modeling applied to Mineral Image Retrieval

Jović Mladen<sup>1</sup>      Yutaka Hatakeyama<sup>1</sup>      Stejić Zoran<sup>2</sup>

Stopić Srečko<sup>3</sup>      Seidl Thomas<sup>4</sup>      Oljaca Dragomirka<sup>5</sup>      Kaoru Hirota<sup>1</sup>

<sup>1</sup> Dept. of C.I. and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, G3-49, 4259 Nagatsuta, Midori-ward, Yokohama 226-8502, Japan

<sup>2</sup> Ricoh Company, Ltd., Software R&D Group, 1-1-17 Koishikawa, Bunkyo-ward, Tokyo 112-0002, Japan

<sup>3</sup> Department of Process Metallurgy and Metal Recycling, RWTH Aachen University, Germany

<sup>4</sup> Data Management and Exploration Group, RWTH Aachen University, Germany

<sup>5</sup> Faculty of Medicine, University of Banja Luka, Bosnia and Herzegovina

E-mail: <sup>1</sup>{jovic, hatake, hirota}@hrt.dis.titech.ac.jp, <sup>2</sup>zstejic@fastmail.fm, <sup>3</sup>sstopic@metallurgie.rwth-aachen.de, <sup>4</sup>seidl@informatik.rwth-aachen.de

## Abstract

Three variants of region-based image similarity models employing the Earth Mover's Distance (EMD) function for automated mineral recognition and identification in the field of mineralogy and metallurgy are presented. By allowing for comparison of images with different number of regions, two models are proposed that take advantage of important characteristics of the EMD, that make it particularly suitable for the region-based image similarity modeling. A third model that takes into consideration five different region-based image similarity models and overcomes the limitations of the existing fixed-resolution image similarity models is further proposed. The three proposed variants share the image features but differ in the way the image feature based flow matrix and feature distribution elements of the EMD are modeled. Empirical evaluation of the three variants on four test databases, containing 3,000 images in 30 semantic categories was carried out. The results obtained from the evaluation revealed that EMD cross-resolution image similarity modeling results in optimal retrieval performance, and provide information useful for adopting the EMD into a region based image similarity modeling. The experimental evaluation presented may thus be helpful for improved mineral recognition and identification.

## I. INTRODUCTION

The study of minerals using optical techniques, principally employing a optical microscope is most commonly undertaken to obtain information such as mineral identity, chemical composition, structural state and growth. (Optical) examination of mineral samples is typically required before more sophisticated techniques are employed, and thus optical identification of minerals is routinely done in the study of rocks and mineral deposits. While this information can be obtained by other more sophisticated techniques such as a *X-ray diffraction*, *X-ray fluorescence* or *electron microscopy*, the methods of *optical mineralogy* have an advantage that they are really simple, inexpensive, reliable, reasonably accurate and require small samples [4]. Until now, mineral *identification* and analysis is mostly done *manually* by fixing and collecting several optical properties

of minerals under polarizing microscope [4], [3]. Manual techniques for mineral recognition and (or) classification are mostly *tedious* and *error prone* and moreover *time consuming*.

With an objective to overcome these limitations *automatic mineral identification and retrieval* based on the digital image analysis is desired. Image processing and retrieval techniques based on the image feature analysis are utilized. In terms to get the best method for automatic and more accurate mineral identification [4], [3], by allowing for comparison of images with different number of regions, two models are proposed that take advantage of important characteristics of the Earth Mover's Distance (EMD) function [17], that make it particularly suitable for the region-based image similarity modeling. A third model that takes into consideration five different region-based image similarity models and overcomes the limitations of the existing fixed-resolution image similarity models is further proposed. The three proposed variants share the image features but differ in the way the image feature based flow matrix and feature distribution elements of the EMD are modeled. Thus overcoming both the fixed number of regions pre-determined by the image segmentation algorithm and the limitations of fixed-resolution image similarity models is achieved. The evaluation is carried out on a mineral image test database, containing 3,000 images in 30 semantic categories. In total, over 900,000 queries are executed, based on which the (weighted) precision and average rank are computed.

The remainder of the paper is organized as follows: In II, Image Similarity Modeling works in Mineral Processing are described. Image Sample Characteristics and Retrieval Strategy are presented in III. The experimental results concerning the effectiveness and efficiency of the proposed methods are outlined in IV.

## II. BACKGROUND: IMAGE SIMILARITY MODELING IN MINERAL PROCESSING

Recognizing different mineralogical species is usually realized by means of *optical microscopy*, sometimes supplemented by *electron microscopy* and *image analysis system* [4], [7]. The analyst is usually an *applied mineralogist* with the great experience and "patience" in carrying out determinations. The work is very boring and repetitive, and for this reason there is always the possibility of the introduction of errors in the determinations.

Techniques for image processing and recognition have been successful in recent years due to the low cost of hardware devices for image acquisition and manipulation and an improvement in their technical characteristics (speed, resolution, processing time, etc.) [4], [5] and [6]. The possibility of using real color images, characterised by triplet of information (three components: RGB), acquired optically, opens a new chapter in the field of process mineralogy. Such procedures in fact have lower costs than systems based on electron beam instruments. Thus a significant aid can be provided by image analysis and the related built-in recognition algorithms. The advantage of these procedures is linked to the fact that they are inexpensive and allow the acquisition, the processing and storing of a large number of images, increasing the significance of the investigation results and *replacing* the experts in the analysis.

As already mentioned, different research groups have tried to techniques for image processing and recognition in mineral recognition purposes. Besides [4], [5] and [6] similar techniques were also developed out and used by Imperial College group [8], CSIRO [9] and CANMET [10]. However, the limitation of these methods are mainly based on image processing techniques where the signal source was not represented by a wavelength distribution (grey-level histogram) in the visible spectrum, but by a source X-rays. Thus it might be rather difficult to arrange and utilize them outside a high technological-scientific environment. In other cases the image processing techniques are based on the black and white images only [13].

The problems of image processing and recognition techniques for mineral recognition purposes are addressed by modeling image similarity by allowing for comparison of images with different number of regions. In such a way , two models are

proposed that take advantage of important characteristics of the Earth Mover’s Distance (EMD) function [17], that make it particularly suitable for the region-based image similarity modeling. A third model that takes into consideration five different region-based image similarity models and overcomes the limitations of the existing fixed-resolution image similarity models is further proposed.

### III. IMAGE SAMPLE CHARACTERISTICS AND RETRIEVAL STRATEGY

As a common approach, the mineral images are transformed into high-dimensional feature vectors and the similarity is measured in terms of vicinity in the feature space. Such a system architecture is shown on Fig. 2. Although the Euclidean

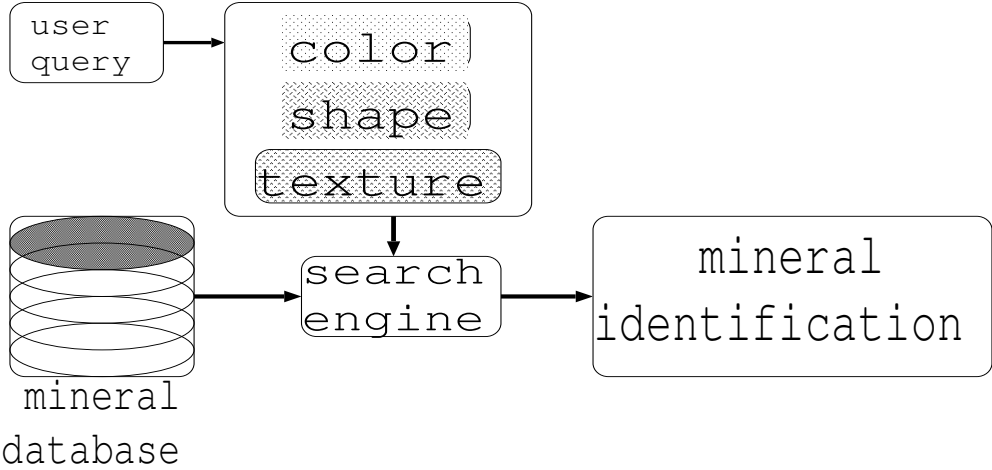


Fig. 1. Proposed system Architecture

distance is a very common distance function for the high-dimensional feature vectors, such as a feature histograms, it exhibits sever limitations with respect to similarity measurements. In particular, the individual components of the feature vectors which correspond to the dimension of the feature space are assumed to be independent of each other [11]. No relationships of the components such as substitutability and compensability may be regarded. To overcome this limitation, image similarity is measured in terms of vicinity in the feature space by means of the Earth Movers Distance function, introduced by Rubner et al. [17]. Accordingly, differences in the three model variants arise from the way in which the image feature distributions [17]  $(X, w) \in \mathbb{R}^{K \times m}$  and  $(Y, u) \in \mathbb{R}^{K \times n}$  are modeled and incorporated into the EMD region-based image similarity modeling.

#### A. Sample Preparation and Data Acquisition

3000 mineral JPEG images of different world-class specimen photos, from the Photo CD collection [2] have been used in the experimental evaluation. Mineral images are 24 bit color depth and  $1024 \times 768$  pixels high-quality of high resolution JPEG standard. Some of the database samples are shown in Fig. 1. Mineral identification from [2] is according to the mineral classification given in the Table I and has been given at [1]. According to [1], the identification key is based on simple mineralogical tests such as luster, hardness, color and physical description for the most common minerals an individual is likely to encounter. Minerals are ordered by their physical and optical properties. The experimental evaluation is done on mineral species according to their color property (cf. Table I). This provides the **ground truth** necessary for the experimental evaluation.



Fig. 2. JPEG images of different world-class specimen photos from Photo CD collection [2]

TABLE I  
MINERALS BY PHYSICAL AND OPTICAL PROPERTIES TABLE CLASSIFIED AS IN [1]

| Physical Properties                          |                              | Optical Properties      |                  |        |
|--|------------------------------|-------------------------|------------------|--------|
| Hardness, Streak, and Luster                 | Cleavage and Fracture        | Density                 | Refractive Index | Color  |
| Metallic by Hardness and Streak              | With No Cleavage             | Metallic by Density     | Isotropic        | Black  |
| Non-Metallic by Hardness and Streak          | With One Cleavage            | Non-Metallic by Density | Uniaxial         | Blue   |
| With a Colored Streak sorted by Streak Color | With Two Cleavages           |                         | Biaxial          | ...    |
|  | With Three or more Cleavages |                         | Opaque           | Yellow |

### B. Mineral Image Feature Representation

**Color** features of mineral images are represented by RGB color histograms, resulting in the 125-D feature vector. Similarity of the image regions is modeled in terms of similarity distributions of color distributions.

**Shape** features of mineral images are represented by the *edge direction histogram* [14], resulting in the 8-D feature vector. Similarity of the image regions is modeled in terms of *similarity distributions of edge directions*.

**Texture** features of mineral images are represented by the texture neighborhood [15], resulting in an 8-D feature vector. Similarity of the image regions is modeled in terms of *similarity distribution of pixels brightness*.

### C. Region-based Image Similarity Modeling

Each sampled image is uniformly partitioned into regions (i.e., rectangular blocks of equal size) at five different resolutions,  $1 \times 1$ ,  $2 \times 2$ ,  $3 \times 3$ ,  $4 \times 4$  and  $5 \times 5$  (resolution  $\in \{1, 2, 3, 4, 5\}$ ). Accordingly, depending on the resolution, the number of regions per image for the above models is between 1 and 25 (number of regions  $\in \{1, 4, 9, 16, 25\}$ ). Each resolution corresponds to a different image similarity model for each of the two EMD region-based model variants. More precisely, image similarity is based on *how* to express the user's similarity criteria represented by means of *low-level image feature representation* (e.g. *color, shape, texture etc.*) and *image feature similarities* (distance function, e.g. *EMD,  $L_1$ ,  $L_2$  etc.*).

### D. Region-based Image Similarity Modeling Variants

1) *Image Area Model*: When incorporating image feature distribution into EMD region-based image similarity modeling :

- an image is described as the *set* of uniformly partitioned image regions, where each region is described by *three finite discrete distributions* of color, shape and texture features.
- for each image region, an element  $x_i$ , ( $1 \leq i \leq 25$ ,  $x \in \{\text{color, shape, texture}\}$ ) belonging to a vector  $x$ , as a component of the histogram employed is equivalent to the "feature" defined in [17].
- for each image region,  $w_i$  ( $\forall i, 1 \leq i \leq 25, w_i \in [0, 1]$ ) corresponds to a *percentage of the relative area of an image region*, covered by the extracted feature vector  $x$ . It is equivalent to the "weight" defined in [17]. As all image regions

cover the same percentage of the overall image area (in any resolution) and there are resolution  $\times$  resolution image regions, all  $w_i$  values are defined by :

$$w_i \equiv \frac{1}{\text{total number of image regions}}, \sum_{i=1}^m w_i = \sum_{i=1}^m \text{region area}_i = 1, \forall i \quad (1)$$

representing the *uniform* weight's distribution of the image region areas within an image.

- thus, an *element* describing the *image region* appears to be the feature vector (*concatenation* of the color, shape and texture descriptors, feature vectors) of one region.

$$\text{element}_i \equiv \left( \underbrace{\text{color}_{\text{region}} + \text{shape}_{\text{region}} + \text{texture}_{\text{region}}}_{x_i \in X}, \underbrace{\text{region area}_{\text{region}}}_{w_i \in W} \right), \quad (2)$$

,  $\text{region} \in \{ 1, \dots, 25 \}$ .

- the set of the vectors corresponding to *all* the regions of an image *image* is equivalent to the *set* :

$$\text{set}_{\text{image}} \equiv (X, w) \in \mathbb{R}^{K \times m} \equiv \{\text{element}_i\} \quad (3)$$

,  $i \in \{ 1, \dots, 25 \}$ .

- the distance between two images is the *single* EMD between two sets associated to the query image  $q$ ,  $\text{set}_q$  and a database image  $i$ ,  $\text{set}_i$ .

$$\text{EMD}(q, i) \equiv \text{EMD}(\text{set}_q, \text{set}_i) \equiv \text{EMD}(\{\text{element}_q\}, \{\text{element}_i\}) \quad (4)$$

When computing the EMD, regions having *different positions* in the two images can be matched. This is not the case if a  $L_p$  distance functions are used between the long vectors grouping the feature vectors of all the regions in an image.

2) *Range Model*: When incorporating image feature distribution into EMD region-based image similarity modeling, the following definitions are applied:

- an image is described as the *set* of uniformly partitioned image regions, where each region is described by *three finite discrete distributions* of color, shape and texture features.
- for each image region, an element  $w_i$ , ( $1 \leq i \leq 25$ ,  $x \in \{\text{color, shape, texture}\}$ ) belonging to a vector  $x$ , as a component of the histogram employed is equivalent to the "weight" defined in [17].
- for each image region, an element  $x_i$ , ( $1 \leq i \leq 25$ ,  $x \in \{\text{color, shape, texture}\}$ ) is equivalent to the "feature" defined in [17].
- thus, an *element* describing the *image region* appears to be a component of the feature vector of either the color, shape or texture descriptor of one region and a component of the vector

$$\text{element}_i \equiv (x_i, w_i), \quad (5)$$

.

- the a vector corresponding to a *single* image region of an image *image* is equivalent to the *set* :

$$\text{set}_{\text{region}} \equiv (X, w) \in \mathbb{R}^{K \times m} \equiv \{\text{element}_i\} \quad (6)$$

,  $i \in \{ 1, \dots, 25 \}$ .

- for every region, distance between regions corresponding to the query image  $q$  and a database image  $i$  having *same* position is calculated as the *average* of the three EMD distances computed for three discrete distributions. This implies

that for each region and its feature, there is a single EMD calculated for each image feature.

$$\text{EMD}_{\text{region}}(q, i) = \frac{1}{3} (\text{EMD}_{\text{color}}(q, i) + \text{EMD}_{\text{shape}}(q, i) + \text{EMD}_{\text{texture}}(q, i)) \quad (7)$$

- the overall image distance between query image  $q$  and a database image  $i$  is calculated as the *average* of EMD distances computed on the regions having same position.

$$\text{EMD}_{\text{Range}}(q, i) = \frac{1}{R} \sum_{\text{region}=1}^R \text{EMD}_{\text{region}}(q, i), \quad \forall R \in \{1, 4, 9, 16, 25\}. \quad (8)$$

where  $R$  denotes number of segmented image regions. The reason for this is that the arithmetic mean (or more generally a subset of a class of the averaging operator) is employed by the most the most image similarity models.

Thus, in contrast to IAM or Cross-resolution model but similar to  $L_p$  distance functions, when computing the EMD, only regions having *same positions* in the two images can be matched.

### E. Cross-Resolution Image Similarity Modeling Model

From the viewpoint of the number of regions used to represent an image content when approximating user's similarity criteria, most of the image similarity models [12], [16], [18], [19] are *fixed* ones, with a respect to a finite number of regions (objects) produced by different segmentation techniques. The limitation of such *fixed* image similarity models is that the existing segmentation algorithms are inaccurate, i.e., the resulting regions are roughly, if at all, corresponding to the objects appearing in the image. To adequately combine the characteristics of different segmentations (expressed as a *single* image similarity model) such that negative peculiarities of the individual segmentations can be masked by others, a *cross-resolution* model as a special variant is proposed which takes into account all five region-based image similarity models simultaneously when computing the image similarity. Initially, each of the five region-based image similarity models represents an image when querying the images within a database on its own as finite number of regions (objects). Next, a *multi* image representation expressed as a set of *multi* image similarity models, each of which is in a different resolution is obtained as a mixture of **all** regions from all five region-based image similarity models. All regions from five fixed-resolution region-based image similarity models are represented as equally important. A *single EMD* computation is utilized on all regions to find out the distance between a set of regions providing an overall similarity between two images. According to its definition, the match between two images is obtained by mixing all regions at *all* different levels. Regions from each fixed-resolution are treated as a *weighted point sets*. The way in which cross-resolution is modeled is as following :

- an image is described as the *set* of **all** uniformly partitioned image regions, provided by five uniform image partitionings. Each region is described by *three finite discrete distributions* of color, shape and texture features.
- for every image region in a *fixed resolution*, an element  $x_i$ , ( $1 \leq i \leq 25$ ,  $x \in \{\text{color, shape, texture}\}$ ) belonging to a vector  $x$ , is equivalent to the "feature" defined in [17]. It is modeled in the same way as for IAM model, described in III-D.1.
- for every image region in a *fixed resolution*, an element  $w_i$ , ( $1 \leq i \leq 25$ , is equivalent to the "weight" defined in [17]. It is modeled in the same way as for IAM model, described in III-D.1.
- thus, an element describing the image region in a *single (fixed) resolution* appears to be the feature vector (*concatenation* of the color, shape and texture descriptors) of one region, just as modeled for the case of IAM model in III-D.1.
- let a term *Cross-Resolution Set* denote the set of the vectors corresponding to *all* the image regions in a *fixed* resolutions:

$$\text{Cross-Resolution Set} \equiv \underbrace{\{(\text{element}_i, \text{region area}_i)\}}_{\text{fixed resolution}}, \quad i \in \{1, \dots, 25\}, \text{resolution} \in \{1, \dots, 5\}. \quad (9)$$

- thus each of the associated *Cross-resolution Sets* for both query image  $q$  and a database image  $i$  consists of five *Resolution Sets*, where each *Resolution Set* is a descriptor of an image in a *fixed* resolution.

$$\begin{aligned} \text{Cross-Resolution Set (q)} &\equiv \underbrace{\underbrace{\text{Resolution Set}_1(q), \dots, \text{Resolution Set}_5(q)}_{\text{resolution 1}}}_{\text{cross-resolution(q)}}, \\ \text{Cross-Resolution Set (i)} &\equiv \underbrace{\underbrace{\text{Resolution Set}_1(i), \dots, \text{Resolution Set}_5(i)}_{\text{resolution 1}}}_{\text{cross-resolution(i)}} \end{aligned}$$

- the distance between two images is the *single* EMD between two sets associated to the query image  $q$ , Cross-Resolution Set( $q$ ) and a database image  $i$ , Cross-Resolution Set( $i$ ).

$$\text{EMD (q, i)} \equiv \text{EMD} \left\{ \underbrace{\text{Cross-Resolution Set (q)}}_{\text{cross-resolution(q)}}, \underbrace{\text{Cross-Resolution Set (i)}}_{\text{cross-resolution(i)}} \right\} \quad (10)$$

- when computing the EMD distance between two sets (associated to two images  $q$  and database image  $i$ ), regions at different resolutions can now be matched (though, weights appear to be different for different resolutions).

When computing the EMD, regions having *different positions* in the two images can be matched. This is not the case if a  $L_p$  distance functions are used between the long vectors grouping the feature vectors of all the regions in an image.

#### IV. EXPERIMENTAL EVALUATION

All the experiments are performed on AMD Athlon 64-bit Processor Machine, with 256 MB RAM Memory. Mineral database [1] is used when conducting experiments, containing 3,000 images divided into 30 semantic categories. Test database originate from the well-known mineral image collection [1], used for the evaluation of mineral image retrieval systems. Partitioning of each database into semantic categories is determined by the creators of the database based on the optical (color) characteristics of the minerals [1]. The query image and the set of candidate images are both predefined by the testing database. The semantic categories define the ground truth so that, for a given query image, the relevant images are considered to be those — and only those — that belong to the same semantic category as the query image. This implies that the number of relevant images for a given query image equals the number of images in the category to which that image belongs.

The performance measures are: (1) precision  $-[P.]$ , (2) weighted precision  $-[W.P.]$ , (3) average rank  $-[A.R.]$ . All the performance measures are computed for each query image, based on the given ground truth. Since each image, in each test database, is used as a query, all the performance measures are averaged for each test database (database-wise average performance). Having defined the computation method of the image similarity based on low-level image features, given a query image, all databases images are ranked in ascending order with respect to their distance to the query image. For each query, all the performance measures are computed.

##### A. Cross and Resolution wise comparison to Global EMD Models

As shown in the Table II, it can be seen that, in terms of retrieval effectiveness measured by P., W.P. and A.R. that IAM model generally performs better than the Range model in all resolutions. In the case of the Range model, in most cases, global image representation (resolution  $1 \times 1$ ) provides optimal retrieval performance. It is decreased by EMD region-based modeling. In case of the IAM model, the opposite is observed.

Even though each pair of matching regions employs a single EMD computation, contrary to expectations, the IAM model outperforms the Range model in both global and region-based image similarity modeling. The first reason for the unexpected

TABLE II

RESOLUTION WISE EVALUATION OF THE PROPOSED EMD REGION-BASED MODELS ON MINERAL IMAGE DATABASE: AVERAGE VALUES OF RETRIEVAL PRECISION (P.), WEIGHTED PRECISION (W.P.), AVERAGE RANK (A.R.) ARE GIVEN. EACH VALUE IS AN AVERAGE OF 3,000 QUERIES EVALUATED OVER THE MINERAL IMAGE DATABASE.

| Resolution         | city-block           |                 |             | <i>Euclidean</i>        |                 |             | <i>IAM</i>     |                 |             | <i>Range</i>  |                 |             |
|--------------------|----------------------|-----------------|-------------|-------------------------|-----------------|-------------|----------------|-----------------|-------------|---------------|-----------------|-------------|
|                    | <i>P. [%]</i>        | <i>W.P. [%]</i> | <i>A.R.</i> | <i>P. [%]</i>           | <i>W.P. [%]</i> | <i>A.R.</i> | <i>P. [%]</i>  | <i>W.P. [%]</i> | <i>A.R.</i> | <i>P. [%]</i> | <i>W.P. [%]</i> | <i>A.R.</i> |
| 1 × 1              | 46.35                | 62.72           | 131.62      | 51.60                   | 68.45           | 118.25      | 51.60          | 68.45           | 118.25      | 36.97         | 53.39           | 155.1       |
| 2 × 2              | 44.74                | 60.81           | 137.37      | 49.88                   | 66.29           | 123.74      | 52.59          | 69.39           | 116.17      | 35.11         | 50.66           | 183.25      |
| 3 × 3              | 42.40                | 58.52           | 143.43      | 33.29                   | 64.19           | 130.02      | 53.74          | 70.46           | 114.59      | 34.86         | 50.59           | 165.2       |
| 4 × 4              | 40.4                 | 56.35           | 149.6       | 32.92                   | 43.63           | 136.93      | 54.38          | 71.13           | 113.14      | 33.15         | 48.69           | 170.98      |
| 5 × 5              | 38.43                | 54.22           | 156.55      | 43.76                   | 60.21           | 144.89      | 55.2           | 71.84           | 111.97      | 30.73         | 45.23           | 173.17      |
|                    | <b>Precision [%]</b> |                 |             | <b>W. Precision [%]</b> |                 |             | <b>A. Rank</b> |                 |             |               |                 |             |
| Cross – resolution | <b>62.39</b>         |                 |             | <b>73.58</b>            |                 |             | <b>67.44</b>   |                 |             |               |                 |             |

result is that, in the case of the Range model, the distribution of the extracted image feature values that are used as "weight" is non-uniform. This is contrary to the first model, in which the distribution of such values is uniform. Furthermore, in terms of shape and texture image features, extracted feature values modeled as "weights" show distribution of edge directions (texture neighborhood pixels) over eight quantized ranges. In the case of color, these are the distributions of image pixel values quantized over 125-D ranges. On the other hand, with respect to the first model, modeling "weight" is completely independent of image visual content, as it is not reflected throughout modeling the "weight". Thus capturing underlying image semantics is expected to be better approximated by the Range model than in the case when the "weight" is modeled independently from the actual visual image content, and the overall EMD distance merely depends on the cost matrix predefined by the extracted feature values. However, the experimental results could be explained by the Range model high data dependency (extracted feature methods used) when modeling feature distributions.

When compared to the cross-resolution EMD model, both IAM and Range are outperformed (Table II). The difference between the cross-resolution model and the global IAM and Range models ranges between 10% and 30%. The integration of five region-based image similarity models (each one for different resolution, into the overall image similarity by EMD matching, accounts for better perceptual similarity than the IAM and Range models. Five fixed-resolution region-based image similarity models employed within the cross-resolution model provide additional information that is not available to the IAM and Range models. The higher discriminative power of the image visual content is ordered by the simultaneous use of five resolutions when modeling image similarity.

As seen in Table II, IAM Model performs optimally when compared to city-block distance function both for global and region-based image similarity modeling. The difference in retrieval performance between the IAM model and Euclidean distance is smaller compared to the Range model.

The larger drop in retrieval performance appears for the Range model compared to the Euclidean and city-block distance. In most cases, optimal retrieval performance for the Range model is provided by global image representation.

In terms of the effect of region-based modeling on retrieval efficiency, it is noted that the Euclidean and city-block distances have a tendency to behave between the IAM and Range models on the mineral database. The retrieval performance is slightly improved (decreased) by increasing the number of regions employed. As shown in Table II in a similar fashion to the Range model, the region-based modeling decreases retrieval performance when Euclidean or city-block distance functions are employed.



## V. CONCLUSION

EMD region-based image similarity modeling and its effect on the retrieval performance applied to mineral image similarity modeling are investigated. By complementary modification of the way in which the cost matrix and feature distributions are modeled, two EMD region-based image similarity models are proposed. An additional special *cross-resolution* EMD region-based image similarity model is also proposed. Comparison and analysis of the image retrieval performance of the three EMD region-based image similarity models is performed. Comparison to representative region-based image similarity models employing Euclidean and city-block distance function is also performed.

The evaluation is carried out on a test database, containing 3,000 images in 30 semantic categories. In total, over 900,000 queries are executed, based on which the (weighted) precision and average rank are computed.

The research strategy in terms of the image similarity modeling applied to the selection of similarity criteria seems to be the most important aspect and strongly influences the quality of results derived from an analysis of this kind. When employed within the EMD region-based image similarity modeling, cross-resolution model provides optimal retrieval performance on mineral database. The difference in the average retrieval precision compared to the two other proposed EMD region-based image similarity models is between 10% and 30%. Contrary to expectations, though possessing the highest computation complexity, the Range model shows the worst retrieval performance. The results of the evaluation suggest that, when modeling region-based similarity modeling within an EMD framework, in terms of the low-level image features, the discretization of the point feature distribution provides better retrieval performance than when low-level image features are modeled as point weights. When modeling image similarity, by allowing different similarity criteria with a different number of region-based similarity models, EMD cross-resolution modeling provides optimal retrieval performance. Such experimental evaluation might be helpful for further understanding of EMD region-based image similarity modeling in various industrial domains, such as mineral species recognition and identification.

## REFERENCES

- [1] Mineral of the Month Club, <http://webmineral.com/specimens.shtml>
- [2] <http://www.excaliburmineral.com/cdintro.htm>
- [3] Pong T.C., Haralick R.M., Craig J.R., Yoon R.H., Choi W.Z., The application of image analysis techniques to mineral processing, in *Pattern Recognition Letters*, Vol. 2(2), 1983, pp. 117–123
- [4] Bonifazi G., Digital multispectral techniques and automated image analysis procedures for industrial ore modelling, in *Minerals Engineering*, Vol. 8(7), 1995., pp. 779–794
- [5] Yan B., McJunkin T.R., Stoner D.L., Scott J.R., Validation of fuzzy logic method for automated mass spectral classification for mineral imaging, in *Applied Surface Science*, Vol. 253(4), 2006., pp. 2011–2017
- [6] Gallagher M., Deacon P., Neural Networks and the Classification of Mineral Samples using X-Ray Spectra in *Proc. of the 9th International Conference on Neural Information Processing, ICONIP 02*. Vol. 5., 2002.
- [7] Wang W., Li L., Image Analysis and Computer Vision for Mineral Froth, in *Proc. of the IEEE International Conference on Mechatronics & Automation*, , Canada, 2005.
- [8] Jones, M.P., Designing an X-ray image analyser for measuring mineralogical data, XIV Int. Min. Proc. Cong., Toronto, VIII-4.1 -4.17, 1982.
- [9] Miller, P.R., Reid, A.F., Zuiderwyk, M.A., QEM\*SEM image analysis in the determination of modal essays, mineral association and mineral liberation, XIV Int. Min. Proc. Cong., Toronto, VIII-3.1 -3.20, 1982.
- [10] Petruk, W., The capabilities of the microprobe Kotron image analysis system: application to mineral analysis, in *Scanning microscopy*, Academic Press, New York, 1973.
- [11] Ankerst M., Kastenmueller G., Kriegel, H.P., Seidl, T., Nearest Neighbor Classification in 3D Protein Databases, in *Proc. of 7th Intl. Conf. on Intelligent Systems for Molecular Biology*, (ISMB99), 1999.
- [12] Jing F., Li M., Zhang H.J., Zhang B., An efficient and effective region-based image retrieval framework in: *IEEE Transactions on Image Processing*, Vol. 13(5), 2004., pp. :699 - 709

- [13] Mengko, T.R., Susilowati, Y., Mengko, R., Leksono, B.E., Digital image processing technique in rock forming minerals identification in: *Circuits and Systems*, 2000. IEEE APCCAS 2000. The 2000 IEEE Asia-Pacific Conference on 4-6 Dec. 2000 Page(s):441 - 444
- [14] Brandt, S., Laaksonen, J., Oja, E., 2000. Statistical shape features in content-based image retrieval. In: *Proc. of 15th Int. Conf. on Pattern Recognition (ICPR-2000) Vol.2*, Barcelona, Spain, pp. 1066-1069.
- [15] Laaksonen, J., Oja, E., Koskela, M., Brandt, S., 2000. Analyzing low-level visual features using content-based image retrieval, In: *Proc. 7th Int. Conf. on Neural Information Processing (ICONIP'00)*, Taejon, Korea, pp. 1333–1338.
- [16] Natsev, A., Rastogi, R., Shim, K., 1999. WALRUS: A similarity retrieval algorithm for image databases, In: *Proc. of ACM SIGMOD Int. Conf. on Management of Data*, Philadelphia, USA, pp. 395–406.
- [17] Rubner, Y., Tomasi, C., Guibas, J.L., 2000. The Earth Mover's Distance as a Metric for Image Retrieval, *International Journal of Computer Vision* 40(2), 99–121.
- [18] Stejić, Z., Takama, Y., Hirota, K., 2003. Comprehensive Comparison of Region-Based Image Similarity Models, *FQAS 2004.*, 390–403.
- [19] Wang, J. Z., Lia J., Wiederhold, G., 2001. SIMPLIcity: Semantics-sensitive Integrated Matching for Picture Libraries, *IEEE Trans. Pattern Analysis and Machine Intelligence* 23(9), 947–963.