1-1-2014

Adaptive and robust feature selection for low bitrate mobile augmented reality applications

Yi Cao
University of Wollongong, yc833@uowmail.edu.au

Christian H. Ritz
University of Wollongong, critz@uow.edu.au

Raad Raad
University of Wollongong, raad@uow.edu.au

Adaptive and robust feature selection for low bitrate mobile augmented reality applications

Abstract

Mobile augmented reality applications rely on automatically matching a captured visual scene to an image in a database. This is typically achieved by deriving a set of features for the captured image, transmitting them through a network and then matching with features derived for a database of reference images. A fundamental problem is to select as few and robust features as possible such that the matching accuracy is invariant to distortions caused by camera capture whilst minimising the bit rate required for their transmission. In this paper, novel feature selection methods are proposed, based on the entropy of the image content, entropy of extracted features and the Discrete Cosine Transformation (DCT) coefficients. The methods proposed in the descriptor domain and DCT domain achieve better matching accuracy under low bit rate transmission than start-of-the-art peak based feature selection used within the MPEG-7 Compact Descriptor for Visual Search (CDVS). This is verified from image retrieval experiments and results for a realistic dataset with complex real world capturing distortion. Results show that the proposed method can improve the matching accuracy for various detectors and also indicate that the feature selection can not only achieves low bit rate transmission but also results in a higher matching accuracy than using all features when applied to distorted images. Hence, even if all the features can be transmitted in high transmission bandwidth scenarios, feature selection should still be applied to the distorted query image to ensure high matching accuracy.

Keywords

robust, feature, selection, low, bitrate, mobile, augmented, adaptive, reality, applications

Disciplines

Engineering | Science and Technology Studies

Publication Details

Y. Cao, C. Ritz & R. Raad, "Adaptive and robust feature selection for low bitrate mobile augmented reality applications," in Signal Processing and Communication Systems (ICSPCS), 2014 8th International Conference on, 2014, pp. 1-7.

Adaptive and Robust Feature Selection for Low Bitrate Mobile Augmented Reality Applications

Yi Cao, Christian Ritz, Raad Raad
ICT Research Institute/School of Electrical Computer and Telecommunication Engineering
University of Wollongong
Wollongong, Australia
yc833@uowmail.edu.au, critz@uow.edu.au, raad@uow.edu.au

***Abstract*—Mobile augmented reality applications rely on automatically matching a captured visual scene to an image in a database. This is typically achieved by deriving a set of features for the captured image, transmitting them through a network and then matching with features derived for a database of reference images. A fundamental problem is to select as few and robust features as possible such that the matching accuracy is invariant to distortions caused by camera capture whilst minimising the bit rate required for their transmission. In this paper, novel feature selection methods are proposed, based on the entropy of the image content in the keypoint domain, the entropy of the extracted features in the descriptor domain and the Discrete Cosine Transformation (DCT) coefficients in the compressed domain. The methods proposed in the descriptor domain and compressed domain achieve better matching accuracy under low bit rate transmission than start-of-the-art peak based feature selection used within the MPEG-7 Compact Descriptor for Visual Search (CDVS) approach while the method proposed in the keypoint domain achieves comparable performance. This is verified from image retrieval experiments and results for a realistic dataset with complex real world capturing distortion including varying lighting conditions, perspective distortion, foreground and background clutter. Results show that the proposed method can improve the matching accuracy for MSER, ORB and SURF detectors which also indicate that the feature selection can not only achieves low bit rate transmission but also result in a higher matching accuracy than using all features when applied to distorted images. Hence, even all the features can be transmitted to server under high transmission network, the feature selection should still be applied to the distorted query image to ensure high matching accuracy.**

Keywords—Feature selection; matching accuracy; low bit-rate transmission

I. INTRODUCTION

The Mobile Augmented Reality (MAR) applications targeted in this paper enhance a user's experience by linking printed media to digital content such as video, picture gallery or webpage [1]–[3]. When a user hovers over a printed image (e.g. an image in the newspaper or magazine) with mobile device camera, the application processes the captured scenes and generates compact visual information for transmission to an image matching system operating on a server. Content related to the matched image is then streamed back to the user. The system diagram of the whole process is shown in Figure 1. The key technology is to analyse the captured scenes and generate a representative compact description for retrieval is highlighted in Figure 1. This process detects and extracts image local features for input to the image matching system. A local feature comprises a keypoint and a descriptor. A keypoint is detected in the image which indicates the coordinates (i.e. (x,y) location) where a local image region contains significant edge information. Then, a descriptor is extracted in the region around a keypoint which characterises the distinctive structural information (e.g. pixel variation) of that local image region. Normally, a descriptor is a vector in which each dimension represents detailed spatial information. The generated local features, for example Scale-Invariant Feature Transform (SIFT) [4], Oriented Robust Binary feature (ORB) [5], Maximally Stable Extremal Regions (MSER) [6] and Speed Up Robust Feature (SURF) [7], should be compact and robust to the distortion caused by the camera capture. They should also have adequate characteristics to perform similarity visual matching in the remote server for retrieval meanwhile the features should also be efficient (i.e. require a low bitrate) for transmission through a wireless network.

Existing research has been done in an on-going MPEG standardization activity known as Compact Descriptor for Visual Search (CDVS) [8]–[10]. To achieve scalability and low bit-rate transmission, certain bit lengths are considered, for example 512B, 1KB, 2KB [11]. However, due to the richness of the captured image scene (e.g. complex visual objects in a scene), hundreds of features resulting in a far longer bit length than these limited bit lengths normally can be detected, which also includes false features caused by noise, such as foreground and background cluttering, varying lighting distortion and perspective distortion. This increases the difficulty of extracting the most significantly discriminative features under these bit lengths within a limited transmission bandwidth. Therefore, an efficient feature selection is desired and such selection criterion is crucial and must be well designed to select the essential local features that eventually can be correctly matched with the target image on the server. This is not only beneficial for low bit-rate transmission but also improving the matching and retrieval accuracy.

To tackle the feature selection problem of MAR applications for low bit rate transmission and high matching

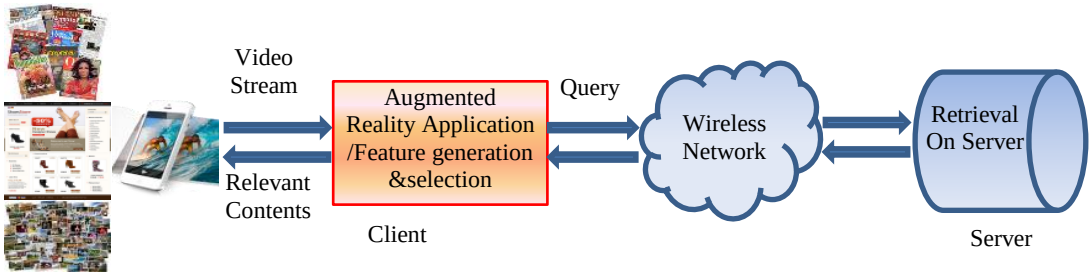

Fig.1. System diagram of targeted MAR applications

accuracy, this paper presents novel feature selection methods based on three metrics: 1) the entropy information of the image content in the keypoint domain; 2) the entropy information of the feature descriptor in the descriptor domain; and 3) the Discrete Cosine Transformation (DCT) coefficients in the compressed domain. The proposed approaches are proven as efficient methods for selecting the most significant and robust features in terms of their ability to result in accurate matching within the system of Figure 1 under different bit-rate constraints and realistic complex capturing distortions. Section II reviews the state-of-art feature selection methods and Section III explains the proposed feature selection methods in detail. Section IV presents retrieval accuracy results for the proposed methods and compares to the state-of-art peak based feature selection. Conclusions are presented in Section V.

II. STATE-OF-ART FEATURE SELECTION METHODS

A critical performance measure of feature selection algorithms is how well their outputs correctly represent the most significant key feature points of an image. It is noted that to achieve such a goal, different keypoint detectors and descriptor extractors can be combined. In this section we review the state-of-art feature selection methods that are relevant to this work.

One of the state-of-the-art solutions in MPEG-7 CDVS is to investigate the relevance of the output parameters of the keypoint detector and the correctly matched feature keypoints [11][12] in the keypoint domain. The output parameters including the Different-of-Gaussian (DOG) response θ_{peak} (denoted as peak in the following paragraphs), scale θ_{scale}, orientation $\theta_{orientation}$, location $\theta_{distance}$ (the distance from the keypoint to the image center) are evaluated individually to investigate the relevance score of these quantities with correctly matched pairs as well as their combination using a probability mass function of correctly matched features learned from dataset. Then, the features are filtered on the basis of sorted relevance scores. The peak of the output of the SIFT detector is superior for identifying the most relevant features compared to other parameters of the output of the SIFT detector, including $\theta_{orientation}, \theta_{scale}, \theta_{distance}$ [12][13]. However, Peak-based Feature Selection (PFS) is constrained to a DoG-based feature detector and is not suitable for other local feature detectors, such as MSER, ORB and SURF. Different feature detectors have different performance in terms of processing speed and matching accuracy. For example, although ORB and SURF are less accurate than SIFT [14][15], their detection time are much faster than SIFT which is desirable for fast processing time on the client side for targeted applications [5][7]. The remaining question is how to improve the matching accuracy for these detectors. Hence, it is desirable to find a generic parameter which can be derived for any feature detector whilst maximizing matching accuracy under low bit-rate feature transmission scenarios.

Alternative feature selection methods take advantage of the underlying discriminating geometric information in the descriptor domain to perform a self-matching method between the original captured image and artificially affine transformed captured image (out-of-plane rotation, flipped) and then chooses the top M matched features [16]–[18]. This method requires a doubling of the feature detection, feature matching and geometric verification processing steps as well as additional image manipulation on the client side, which consumes more computational resource and battery power. It is also difficult to accurately determine the thresholds used in these feature matching and geometric verification stages used for feature selection for a wide variety of images. Hence, this paper proposes an alternative approach that avoids this doubling of the process and additional image manipulation.

III. PROPOSED FEATURE SELECTION METHOD

An efficient method should be found to utilize not only the output parameters of feature detector but also the implicit information embedded in the local image patch and feature descriptor to select the most significant and robust features according to varying low bit-rate requirement. In this section, the problem of selecting the key features for matching a captured frame to a reference image is firstly formatted and secondly the proposed feature selection method based on the entropy information of the image content in the keypoint domain and SIFT features in the descriptor domain as well as the DCT coefficients in the compressed domain are proposed.

A. Problem Formatting

The problem of selecting the key features of a captured image to match an image in a remote database containing N candidate images can be formulated as follows (the images used in this work are grayscale images for content-based image matching):

1) Assuming the captured image is represented by the feature set $X = (x_1, x_2 \dots x_L), x_i \epsilon R^m$; the N candidates in the database are represented by the feature set $\{Y_1, Y_2 \dots Y_N\}$, $Y_i = (y_1, y_2 \dots y_K), y_j \epsilon R^m$;

2) Assuming that the probabilities of the captured image being correctly matched to each candidate are $H = (h_1, h_2 \dots h_N)$; $h_i = f(X, Y_i)$ where $f(\cdot)$ measures the

similarity between X and Y_i;

3) If the *m-th* candidate in the database is corresponding to X, the objective is to find a proper metric θ to select the key features that makes $P(X|h_m) > P(X|h_i), where\ i\epsilon(1,N), i \neq m$, for example, $\theta_{peak}, \theta_{orientation}, \theta_{scale}, \theta_{distance}$ in [12].

B. Proposed feature selection methods

Three metrics in different domains are considered for feature selection in this work: 1) Keypoint domain using Local Patch Entropy (LPE); 2) Descriptor domain using Descriptor Entropy (DE); 3) Compressed domain using DCT coefficients of a local patch around a keypoint. The definitions of these metrics are described in this subsection. Additionally, to study the generality and applicability of the proposed methods, different combinations of detector and descriptor are employed for investigation, including MSER detector, ORB detector, SURF detector, DOG detector and SIFT descriptor.

1) Local Patch Entropy

The local entropy is used to determine the local complexity of an image [19]. Intuitively, the local entropy is an efficient metric to select the features. After the feature detection, given a detected feature point *x*, a local neighborhood $\mathrm{R_x}$ around that feature point which takes on pixel values $\{\mathrm{r}_1, \dots, \mathrm{r}_\mathrm{m}\}$, local patch entropy can be calculated as:

$$\theta_{LPE} = -\sum_i P_{R_x}(r_i) log_2 P_{R_x} \quad (1)$$

where $P_{R_x}(r_i)$ is the probability of r_i on the histogram of 0~255 using 256 bin as the grayscale image is used in this work. Thus, each detected feature point *x* can be assigned with a θ_{LPE}. The probability of θ_{LPE} of features being correctly matched can be learned from the dataset and then such probability can be used to rank features for selection.

2) Descriptor Entropy

The local feature descriptor normally encapsulates certain high level characteristics extracted from pixel values. For example, the SIFT descriptor encapsulates the gradient and orientation information around the keypoint [4]. The assumption is that the more entropy the descriptor has, the more distinctive information is encapsulated in the descriptor thus the more important the descriptor is. Given a detected feature point *x* and a corresponding n-dimensional descriptor, $D_x \epsilon R^n$ takes a value on each dimension $\{d_1, \dots, d_n\}$ and encapsulates the high level information around a keypoint. The descriptor entropy can be calculated as:

$$\theta_{DE} = -\sum_i P_{D_x}(d_i) log_2 P_{D_x} \quad (2)$$

where $P_{D_x}(d_i)$ is the probability of d_i on the histogram of 0~255 using 256 bin as SIFT descriptor is used in this work and each dimension of SIFT feature is represented by 8 bit. Therefore, each detected feature point *x* can be assigned with a θ_{DE} computed from the corresponding descriptor. After learning the probability of θ_{DE} of features being correctly matched from the dataset, the features can be ranked for selection based on θ_{DE}.

3) DCT coefficients of a local patch around keypoint

The DCT coefficients have been widely used for compressed domain retrieval and it is known that the DC component and first two AC coefficients contain the main structure information of the image [20]–[22]. In this work, the DCT coefficients are employed as a proper metric for feature selection. Given a detected feature point *x*, a 16*16 local image patch around the keypoint (as the region of a SIFT descriptor is 16*16), 16*16 2D-DCT transformation is applied in the local patch to calculate the DCT coefficients θ_{DCT}:

$$\theta_{DCT}(u,v) = \alpha_u \alpha_v \sum_{x=0}^{M-1} \sum_{y=0}^{n-1} f(x,y) \cos\frac{\pi(2x+1)u}{2M} \cos\frac{\pi(2y+1)v}{2N} \quad (3)$$

where

$$\alpha_u = \begin{cases} \frac{1}{\sqrt{M}}, u = 0 \\ \sqrt{\frac{2}{M}}, 1 \le u \le M-1 \end{cases}; \alpha_v = \begin{cases} \frac{1}{\sqrt{N}}, v = 0 \\ \sqrt{\frac{2}{N}}, 1 \le v \le N-1 \end{cases}$$

Here, *M*=*N*=16. In this work, we mainly consider the following DCT coefficients θ_{DCT} : $\theta_{DC} = \theta_{DCT}(0,0)$; $\theta_{AC1} = \theta_{DCT}(0,1)$; $\theta_{AC2} = \theta_{DCT}(1,0)$ as these components contain the main structural information of the local patch compared to higher frequency AC coefficients [20][22] . Therefore, each detected feature point *x* can be assigned with a series of θ_{DCT}. The probabilities of DCT coefficients of the local patch around correctly matched features are computed from the dataset and used for ranking features for selection.

4) Learning the probabilities for feature selection

The key stage of the proposed method is to learn the probabilities of proposed feature selection metrics to measure how well a feature can be correctly matched from the dataset which is denoted as 'matchability' of a feature. For all the features extracted from the images of the dataset, the proposed metrics are calculated using Equation (1) ~ (3) for each feature, respectively. Then the correctly matched features are learned from the supervised pair-wise image matching. For a specific metric, for example θ_{DE}, it is divided into *N* bins. The histogram of all the features for $\theta_{DE}$is calculated and denoted as $h(\theta_{DE\_all})$ while the histogram of correctly matched features for θ_{DE} is denoted as $h(\theta_{DE\_match})$. The 'matchability' of features according to θ_{DE} is defined as:

$$Matchability(\theta_{DE}) = \frac{h(\theta_{\mathrm{DE}\_match})}{h(\theta_{\mathrm{DE}\_all})} \quad (4)$$

The 'matchability' of θ_{LPE} and θ_{DCT} is calculated using Equation 4 as well. The detailed procedure for learning 'matchability' is explained in Section IV-B following the description of dataset.

IV. EXPERIMENT OF USING FEATURE SELECTION FOR LOW BIT-RATE RETRIEVAL

A. Experimental dataset

To test the effectiveness of the proposed method for the low bit-rate mobile augmented reality applications targeted in this work, the printed media images from MVS dataset [23] are used. This dataset corresponds to the main printed media in the CDVS dataset [8] and contains more than 1200 camera-phone captured different types of print images including CD covers, DVD covers and book covers. These images are denoted as query images. The data set has several key characteristics that reflect realistic situations: rigid objects, widely varying lighting conditions, perspective distortion, foreground and background clutter, and query data collected from heterogeneous low and high-end camera phones. The ground-truth images are also available and used for learning the probabilities by performing pairwise matching. These

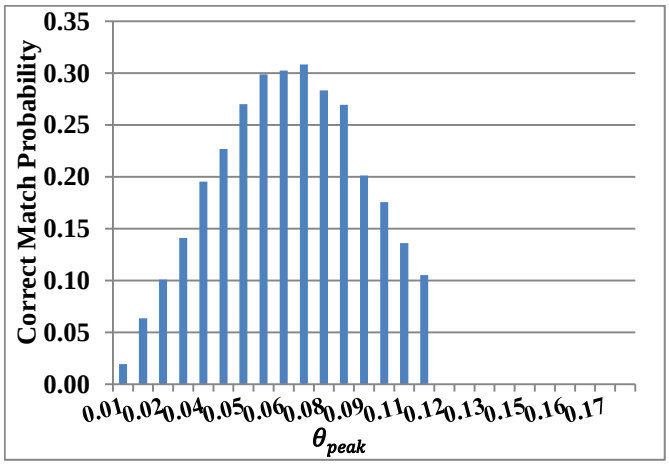

Fig. 2. Probability of correctly matched feature pairs across the whole dataset vs. θ_{peak}

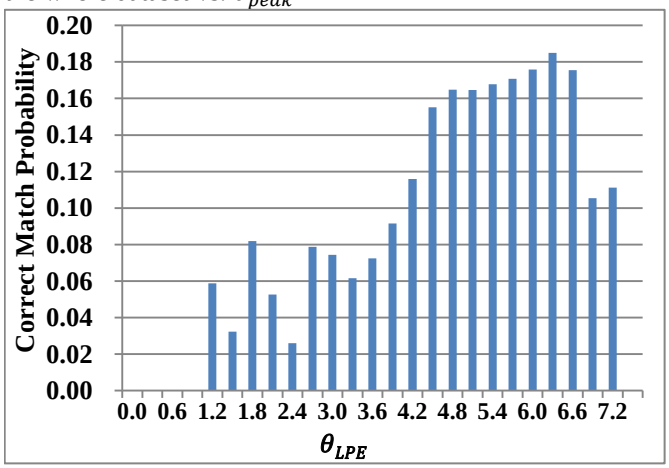

Fig. 3. Probability of correctly matched feature pairs across the whole dataset vs. θ_{LPE}

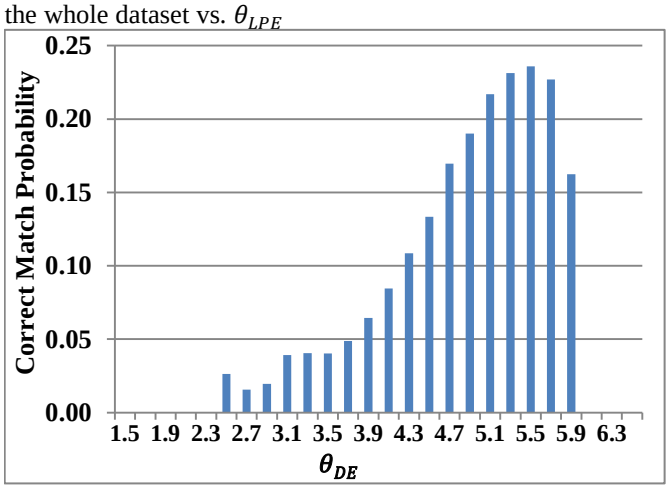

Fig. 4. Probability of correctly matched feature pairs across the whole dataset vs. θ_{DE}

ground-truth images are denoted as reference images.

B. The methodology of Learning 'matchability'

To learn the 'matchability', the image matching pair list of each query image and reference image is established according to the provided ground-truth images. Both images in a pair depict the same object. Learning the probabilities of proposed metrics proceeds automatically using the image matching pair list and is carried out on image pairs. The peak value of the DOG detector in MGEG-7 CDVS is also used for comparison. Each image pair undergoes the following process:

1). Detect keypoints and extract SIFT descriptors both from

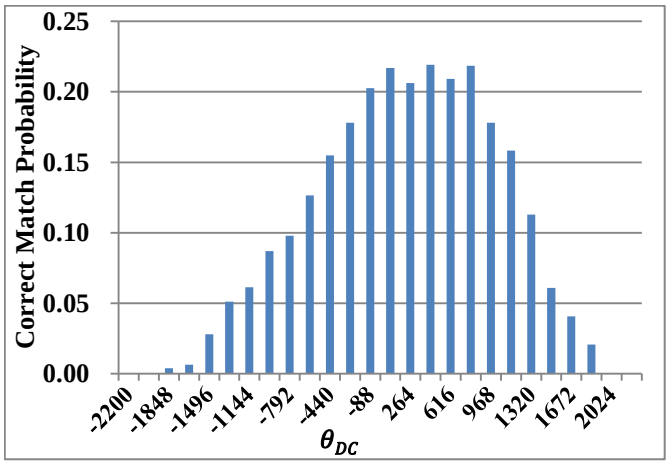

Fig. 5. Probability of correctly matched feature pairs across the whole dataset vs. θ_{DC}

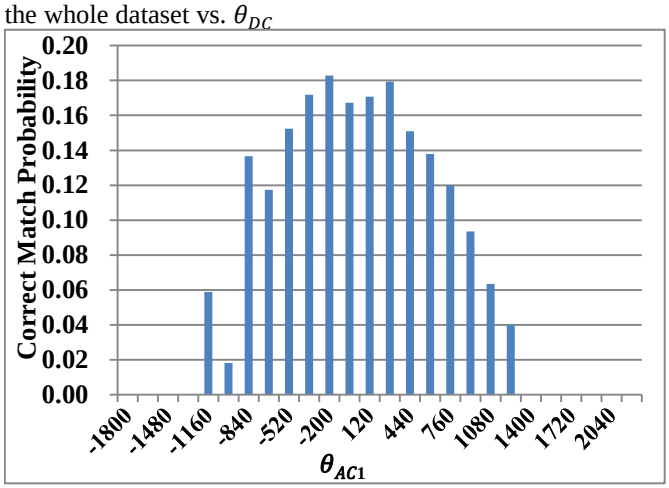

Fig. 6. Probability of correctly matched feature pairs across the whole dataset vs. θ_{AC1}

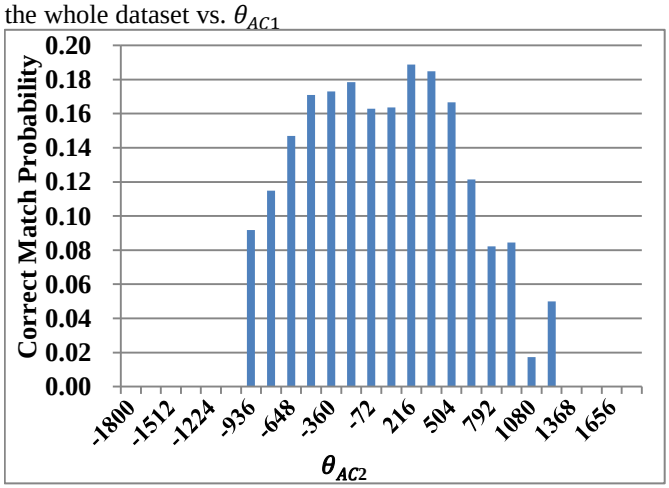

Fig. 7. Probability of correctly matched feature pairs across the whole dataset vs. θ_{AC2}

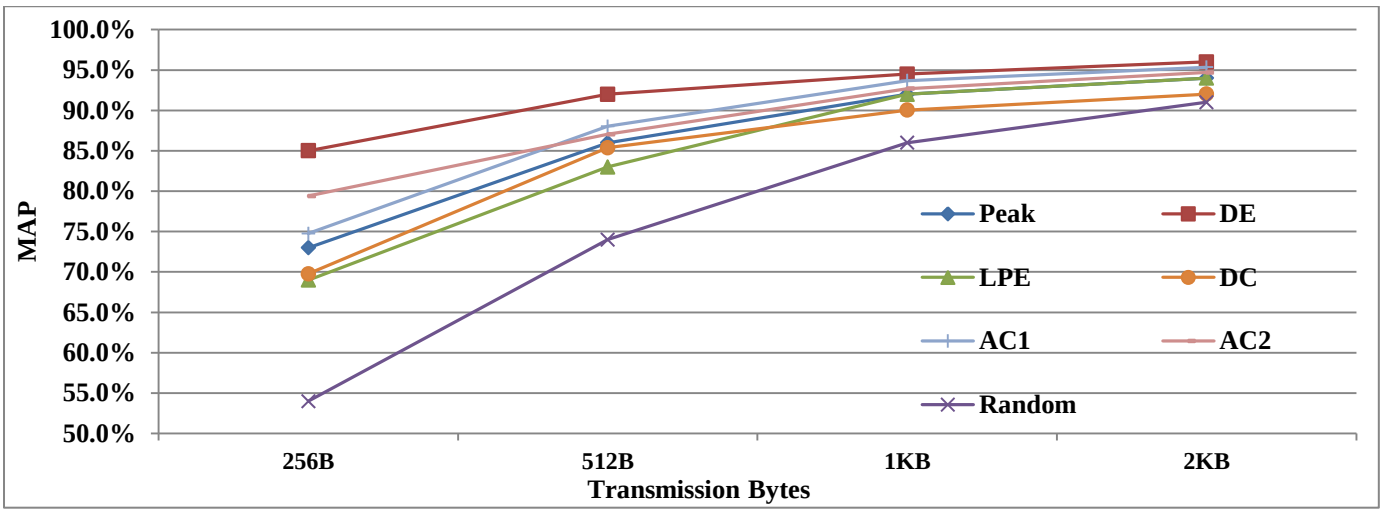

Fig. 8. The retrieval performance of proposed methods compared with peak-based and random feature selection method under varying low bitrate.

query and reference images in the database. For each feature, the peak value, local entropy, descriptor entropy and DCT coefficients $\{\theta_{peak}, \theta_{LPE}, \theta_{DE}, \theta_{DC}, \theta_{AC1}, \theta_{AC2}\}$ are computed and recorded for each feature using Equation (1) ~ (3);

2). Perform the Nearest Neighbor search (i.e. KNN search where k=1 [24]) within each image pair to find the nearest neighbor for each feature and then perform the cross-check method to select features. This method only returns feature matching pairs (i, j) where the i-th query descriptor from query image is nearest to the j-th descriptor from reference image in the matcher's collection and vice versa [25].

3). Perform Geometric Verification using RANSAC [26] and the remaining features are taken as true positive features;

4). Calculate the probabilities of the true positive features (i.e. correctly matched features) using Equation (4) for $\{\theta_{peak}, \theta_{LPE}, \theta_{DE}, \theta_{DC}, \theta_{AC1}, \theta_{AC2}\}$ individually. The learned probabilities of $\{\theta_{peak}, \theta_{LPE}, \theta_{DE}, \theta_{DC}, \theta_{AC1}, \theta_{AC2}\}$ are shown in Figure 2 ~ Figure 7.

C. *The probability learned from the dataset for feature selection*

From the Figure 2 ~ Figure 7, it is intuitively known that $\{\theta_{peak}, \theta_{LPE}, \theta_{DE}, \theta_{DC}, \theta_{AC1}, \theta_{AC2}\}$ are effective for filtering the features as they all exhibit a certain distinctive distribution. Each detected feature can be assigned a probability according the $\{\theta_{peak}, \theta_{LPE}, \theta_{DE}, \theta_{DC}, \theta_{AC1}, \theta_{AC2}\}$. After assigning the probability to each feature based on these distributions, the features are ranked from high probability to be matched to low probability. The feature sets can be easily filtered on the basis of ranked features using a feature number threshold according to different application requirements in terms of bitrate.

D. *Retrieval experimental result*

To tackle the problem of the targeted application, the proposed methods are applied to the retrieval task under different bitrates to transmit varying number of features. The experimental procedure is as follows:

1) For each query image in the dataset:

(a) Detect and extract the features;

(b) Select the specified number of features using the proposed feature selection methods. This forms the query feature set with the remaining features filtered out;

2) For the reference images in the dataset:

(a) Detect and extract the features for each reference image;

(b) Combine the detected features of each reference imag to set up the training feature set;

(c) Perform KD-tree training to obtain the reference feature search space.

3) For each query feature set:

(a) Perform the nearest neighbor search using KNN (k=1) for each query feature in the trained reference feature search space;

(b) Obtain the first N (N=3) reference images with maximum feature matching pairs(Increasing N did not bring out significantly better retrieval results) ;

(c) Perform cross-check KNN (k=1) search within each chosen reference image to further filter the features;

(d) Apply geometric verification (RANSAC) to find the final true positive feature matching pairs.

(e) Locat the reference image on the basis of the highest number of true positive feature matching pairs;

(f) Declare a correct match using a ground truth file list.

The matching accuracy is evaluated based on the Mean Average Precision (MAP) to judge the retrieval performance [8], [27] under different bitrate:

$$MAP = \frac{1}{Q}\sum_{q=1}^{Q} P(q)$$
$$P(q) = \begin{cases} 1, & the\ matched\ image\ is\ correct \\ 0, & otherwise \end{cases} \quad (5)$$

Q is the number of query images.

For comparison, the retrieval experimental results of using the proposed feature selection methods, the peak-based feature selection in MPEG-7 CDVS and random feature selection for DOG detector and SIFT descriptor are presented in Figure 8. The random feature selection generates a random keypoint index list to choose features. Four different feature number conditions are considered in the experiment 279, 210, 114 and 50 which correspond to 2KB, 1KB, 512B and 256B compressed feature transmission sizes. The first three bit rates are standardized in the MPEG-7 CDVS [10]. The fourth bit rate is also considered in the scenario of a very poor communication condition or processing condition where a very fast transmission is desired (e.g. processing a stream of video frames to repeatedly look for a matching reference image).

From Figure 8, it is evident that DE outperforms the peak-based method, especially at low bit rates. DE achieves a 6% and 12% retrieval performance gain for 512B and 256B respectively. The AC1 and AC2 can also efficiently select the important features which achieve 6% and 2% better performance than peak-based method at 256B, respectively. The LPE and DC are comparable to the peak-based method and the worst performance degradation is only 3% when using 256B. As expected, the random selection method (i.e. randomly choosing a certain number of features without any criteria) degrades the matching accuracy compared to the other methods. For 2KB transmission (i.e. 279 features), the random method still achieves 90% because it selects on average more than 85% of the features generated by the SIFT algorithm. (the total number of detected SIFT feature is determined by the complexity of an image).

To study the generality and applicability of the proposed methods, another three feature detectors for which the peak value is unavailable are employed. These are MSER [6], ORB [5] and SURF [7]. The MAP results of different detectors under different bit rate are shown in Table 1~4. As different detectors result in different matching accuracy, to show the effect of the proposed selection methods, the MAP gain (difference between the MAP results of using selection methods and MAP results without selection) are presented in Figure 9. The positive values in Figure 9 indicates the MAP is improved by employing selection methods compared to the MAP result without selection method while negative values indicates the degradation of MAP. The equation for calculating the MAP gain is defined as:

$$\text{MAP\_Gain} = \frac{MAP_{Selection} - MAP_{noselection}}{MAP_{noselection}} \qquad (6)$$

where $MAP_{Selection}$ is the MAP result using feature selection methods, $MAP_{noselection}$ is the MAP result without feature selection. The legend denotes the used detector and selection method as 'Detector-Selection', for example, using MSER as detector and DE as selection method are referred as MSER-DE.

The MAP results without feature selection methods for MSER, ORB, and SURF are 28%, 42%, 72% as shown in Table 4, respectively which are consistent with the results in [14][15]. The MSER and ORB did not achieve good MAP due to complex distortions in the experimental dataset. However, we are more interested in how the proposed method can improve the MAP result. Figure 9 shows that the proposed feature selection methods improve the matching accuracy for all features. The maximum gains are 28.5% for MSER using θ_{AC2} at 2KB and 67.2% for ORB using θ_{DE} at 512B. For SURF, only the DE method achieves a maximum of 16.7% gain at 512 KB while the other selection methods lead to a negative gain (as much as 22% degradation at 256B for LPE). The main reason for the improvement of matching accuracy is that the false positive features are filtered out which is beneficial to the cross check matching and geometric verification. Hence, to maximize the MAP under distorted query images, it is suggested that the selection method is chosen based on the image feature and transmission bit rate

Table. 1. The MAP results of MSER detectors under different bitrate using different feature selection methods

MSER				
	256B	**512B**	**1KB**	**2KB**
DE	0.3040	0.3040	0.3400	0.3120
PE	0.2920	0.3520	0.3480	0.3440
DC	0.2880	0.3560	0.3360	0.3360
AC1	0.2840	0.3240	0.3360	0.3120
AC2	0.2840	0.3200	0.3320	0.3600

Table. 2. The MAP results of ORB detectors under different bitrate using different feature selection methods

ORB				
	256B	**512B**	**1KB**	**2KB**
DE	0.6262	0.7025	0.6491	0.5804
PE	0.4582	0.4505	0.5575	0.4815
DC	0.5116	0.4276	0.4582	0.4583
AC1	0.4278	0.4964	0.5727	0.4582
AC2	0.4277	0.5269	0.4735	0.4735

Table. 3. The MAP results of SURF detectors under different bitrate using different feature selection methods

SURF				
	256B	**512B**	**1KB**	**2KB**
DE	0.7938	0.8400	0.8262	0.7431
PE	0.5585	0.5908	0.6231	0.6738
DC	0.7154	0.7246	0.6415	0.6554
AC1	0.6738	0.6831	0.7246	0.6692
AC2	0.7108	0.7246	0.7246	0.6508

Table. 4. The MAP results of MSER, ORB, SURF detectors without selection (i.e. all the detected features from query image are used for matching)

MSER	**ORB**	**SURF**
0.28	0.42	0.72

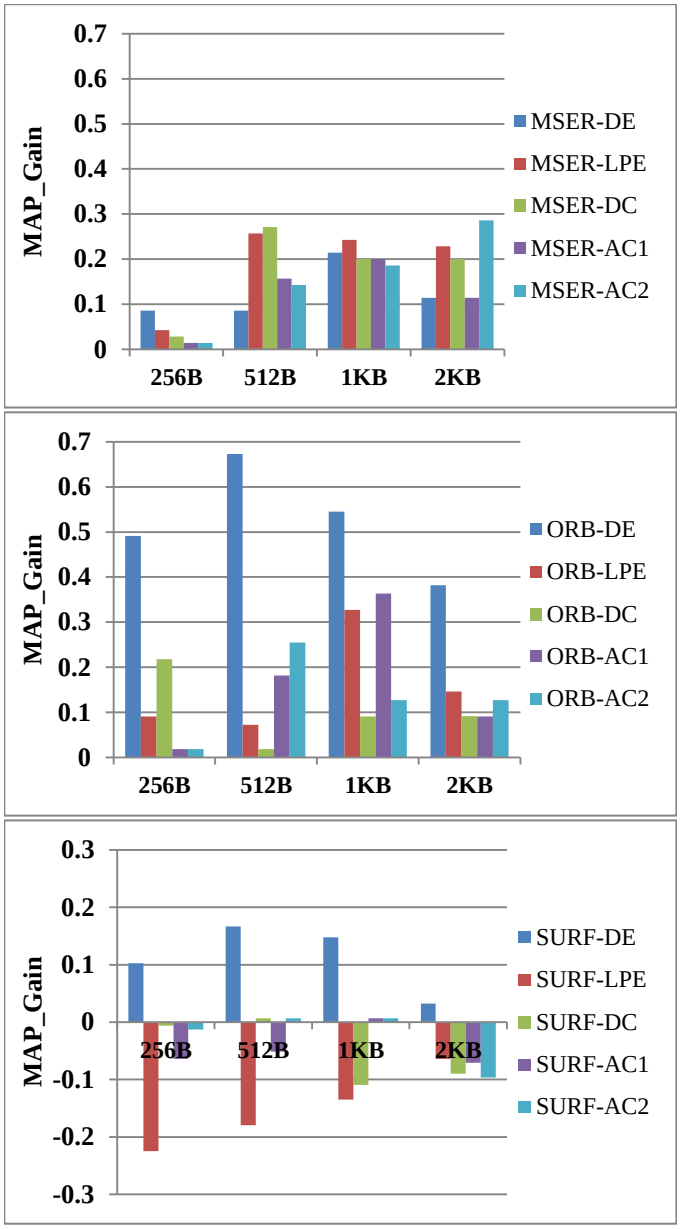

Fig. 9. The MAP improvement results of using different selection methods compared to the method without selection.

being used in the matching system.

V. CONCLUSION

Novel methods for feature selection are proposed in this paper by which a subset of robust detected features in terms of their ability to correctly match a captured image to a reference image can be selected and transmitted at low bitrate to retrieve an augmented multimedia content accurately. The proposed metrics take advantage of the discriminative information embedded in the entropy of the local image patch, entropy of the descriptor and DCT coefficients for feature selection. When compared to start-of-the-art peak based feature selection, the proposed methods based on descriptor entropy and DCT coefficients achieve superior image retrieval performance on a dataset with complex realistic distortions. The proposed methods also improve the matching accuracy of MSER, ORB and SURF detectors which not only prove the generality and applicability of the proposed methods but also indicate that the feature selection should be still applied to the distorted query images to ensure high matching accuracy even all the features can be transmitted to server under high transmission network. The future work may be extended to study the combination of these different metrics for feature selection and the influence of different image types (e.g. rigid object vs. non-rigid object) for feature selection methods.

ACKNOWLEDGMENT

This work is supported by Smart Services CRC, Sydney, Australia.

REFERENCES

[1] S. Davis, E. Cheng, C. Ritz, and I. Burnett, 'Ensuring Quality of Experience for markerless image recognition applied to print media content', in *2012 Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2012, pp. 158 –163.

[2] 'viewa', *viewa*. [Online]. http://viewa.net/.

[3] 'Home | Augmented Reality | Interactive Print', *Layar*. https://www.layar.com/.

[4] D. G. Lowe, 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[5] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, 'ORB: an efficient alternative to SIFT or SURF', in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 2564–2571.

[6] J. Matas, O. Chum, M. Urban, and T. Pajdla, 'Robust wide-baseline stereo from maximally stable extremal regions', *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, 2004.

[7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, 'Speeded-up robust features (SURF)', *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.

[8] ISO/IEC JTC1/SC29/WG11/N12551, 'CDVS, Description of Core Experiments on Compact descriptors for Visual Search'. Feb-2012.

[9] ISO/IEC JTCl/SC29/WG11/N12550, 'Test Model 1: Compact Descriptors for Visual Search'. Feb-2012.

[10] ISO/IEC/JTC1/SC29/WG11/W12929, 'Test Model 3: Compact Descriptor for Visual Search'. Jul-2012.

[11] 'Study Text of ISO/IEC CD 15938-13 Compact Descriptors for Visual Search'.http://mpeg.chiariglione.org/standards/mpeg-7/compact-descriptors-visual-search/study-text-isoiec-cd-15938-13-compact-descriptors.

[12] G. Francini, S. Lepsøy, and M. Balestri, 'Selection of local features for visual search', *Signal Process. Image Commun.*, vol. 28, no. 4, pp. 311–322, Apr. 2013.

[13] K. Lee, S. Lee, S. Na, S. Je, and W.-G. Oh, 'Extensive analysis of feature selection for compact descriptor', in *2013 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision, (FCV)*, 2013, pp. 53–57.

[14] Y. Cao, C. Ritz, and R. Raad, 'Image compression and retrieval for Mobile Visual Search', in *Communications and Information Technologies (ISCIT), 2012 International Symposium on*, 2012, pp. 1027–1032.

[15] O. Miksik and K. Mikolajczyk, 'Evaluation of local detectors and descriptors for fast feature matching', in *2012 21st International Conference on Pattern Recognition (ICPR)*, 2012, pp. 2681–2684.

[16] X. Xin, Z. Li, Z. Ma, and A. K. Katsaggelos, 'Robust feature selection with self-matching score', in *2013 20th IEEE International Conference on Image Processing (ICIP)*, 2013, pp. 4363–4366.

[17] G. Tolias, Y. Kalantidis, and Y. Avrithis, 'SymCity: feature selection by symmetry for large scale image retrieval', in *Proceedings of the 20th ACM international conference on Multimedia*, 2012, pp. 189–198.

[18] ISO/IEC JTC1/SC29/WG11/M23929, 'Reference results of key point reduction'. 99th MPEG Meeting, Sanjose, USA, 2012.

[19] T. Kadir and M. Brady, ‘Saliency, scale and image description’, *Int. J. Comput. Vis.*, vol. 45, no. 2, pp. 83–105, 2001.
[20] C.-W. Ngo, T.-C. Pong, and R. T. Chin, ‘Exploiting image indexing techniques in DCT domain’, *Pattern Recognit.*, vol. 34, no. 9, pp. 1841–1851, Sep. 2001.
[21] F. Arnia, I. Iizuka, M. Fujiyoshi, and H. Kiya, ‘Fast Method for Joint Retrieval and Identification of JPEG Coded Images Based on DCT Sign’, in *IEEE International Conference on Image Processing, 2007. ICIP 2007*, 2007, vol. 2, pp. II – 229–II – 232.
[22] D. Edmundson and G. Schaefer, ‘An overview and evaluation of JPEG compressed domain retrieval techniques’, in *ELMAR, 2012 Proceedings*, 2012, pp. 75 –78.
[23] V. R. Chandrasekhar, D. M. Chen, S. S. Tsai, N.-M. Cheung, H. Chen, G. Takacs, Y. Reznik, R. Vedantham, R. Grzeszczuk, J. Bach, and B. Girod, ‘The Stanford Mobile Visual Search Data Set’, in *Proceedings of the Second Annual ACM Conference on Multimedia Systems*, New York, NY, USA, 2011, pp. 117–122.
[24] M. Muja and D. G. Lowe, ‘Fast approximate nearest neighbors with automatic algorithm configuration’, in *In VISAPP International Conference on Computer Vision Theory and Applications*, 2009, pp. 331–340.
[25] ‘Common Interfaces of Descriptor Matchers’.: http://docs.opencv.org/modules/features2d/doc/common_interfaces_of_descriptor_matchers.html.
[26] M. A. Fischler and R. C. Bolles, ‘Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography’, *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[27] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*, vol. 1. Cambridge University Press Cambridge, 2008.