

Utilization of Textural Features in Video Retrieval System by Hand-writing Sketch

Hiroki Kobayashi¹

¹ School of Systems Information Science, Future University Hakodate, Hokkaido, Japan
e-mail: b1008210@fun.ac.jp

Masashi Toda²

² Center for Multimedia and Information Technologies, Kumamoto University, Kumamoto, Japan
e-mail: toda@cc.kumamoto-u.ac.jp

Abstract— In this paper we propose a video retrieval system using texture features from a hand-drawn sketch. There is currently a lot of video content on the Internet due in part to the development of video-sharing Web sites, and sometimes users are unable to narrow the options down to the desired video because there is simply too much content to choose from. We should be able to solve this problem with a retrieval technique that uses internal movie features. Previously, our laboratory proposed a video retrieval system using three internal movie features and a sketch query. In the current study, our aim is to improve the retrieval precision by adding a new feature derived from texture.

Index Terms— Video Retrieval, Sketch, Texture, Density

I. INTRODUCTION

These days, video content on the Internet is constantly increasing. A basic keyword search is one of the ordinary methods of retrieving the desired video, but due to the growing amount of content it is becoming increasingly difficult to rely on the keyword search method alone.

For example, if we type the keywords “dog ball playing” into GoogleVideo, about 33,100 videos are retrieved. To find our desired video from among all these, we need to check each title and the content or else somehow narrow down the search, usually by adding more keywords. Obviously, it takes a lot of time and effort to check each video, and adding keywords is also problematic because there are simply not enough keywords to adequately describe a user's mental image in terms of screen structure, action, etc. This is a serious issue because screen structure and action are both such important elements in movies. We expect to retrieve user's desired movie by using such features. Video content on the Internet is constantly, and therefore video retrieval systems that use actual video features is going to become more important.

Many video retrieval systems that use video features already exist. One of these is a video retrieval system using hand-drawing sketches proposed by Sekura et. al[1]. In this paper, we describe our own video retrieval system with texture features that is based on the system.

II. RELATED WORK

In this section, we describe other works that are related to our proposed system.

As mentioned above, Sekura et. al[1] have proposed a video retrieval system that uses a hand-drawn sketch query to search. The sketch query is based on the idea of “describing one scene with one sketch”. The system make ranking of movies on the basis of a feature's distance from movement, the shapes of objects, and the shapes of backgrounds. They used handwriting sketch query to search. The sketch query based in method “Describe one scene by one sketch”. The system is lay out movies base on feature distance from movement, shapes of object, shapes of background.

Figure 1 is SEKURA proposed system process flow. The system extracted moving object area in video each frame based optical flow. Moreover, extracted feature shapes of object from moving object area and shape of background from outside moving object area. Movement feature from both area optical flow.

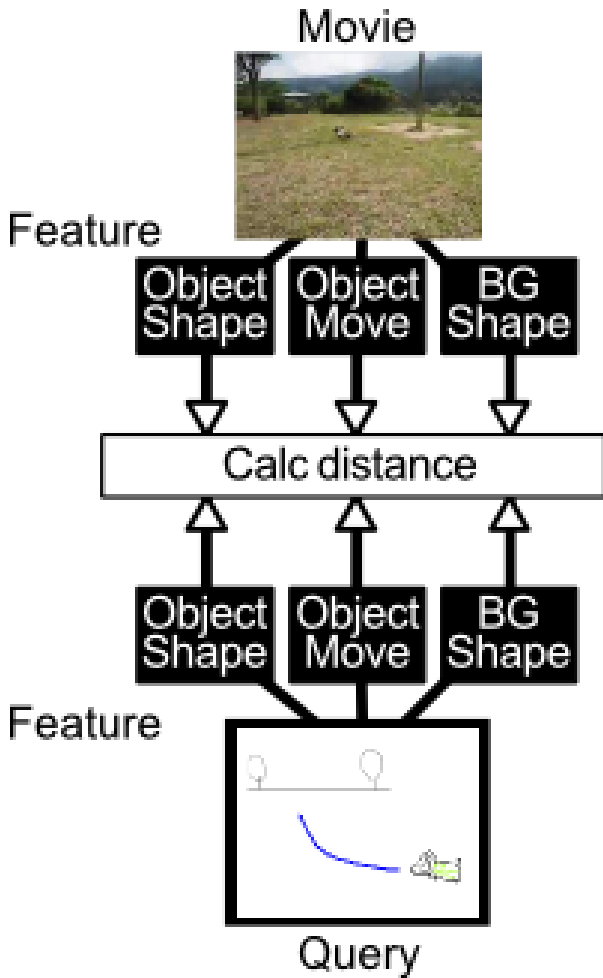
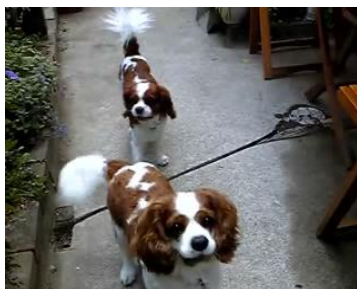
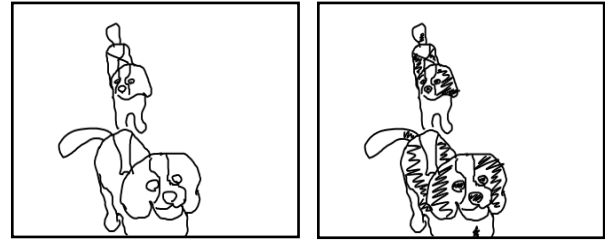


Figure : 1 The flow of Sekura's System

Igarashi et. al[2] have proposed using a hand-drawn sketch as a image search feature. The system is distinct from Sekura's in that they use the sketch's texture information, which here means the rough fill pattern description. In the case of a monochrome sketch, colors and patterns can be deduced from the texture information (Fig. 2). Igarashi's system is focused primarily on texture indication properties, which to viewers appear as dim areas. Therefore, they associate texture indication with brightness in an image.



(a) Sketch-based Image



(b) Non-textured Sketch

(c) Textured Sketch

Figure 2. Texture in Sketch

III. PROPOSED SYSTEM

In this section, we describe the algorithm and features of the proposed system. Its process flow is shown in Fig.3. In terms of process flow, the proposed system is similar to Sekura's, but we have added a function to evaluate texture features, which ultimately changes the search query input method and distance calculation formula.

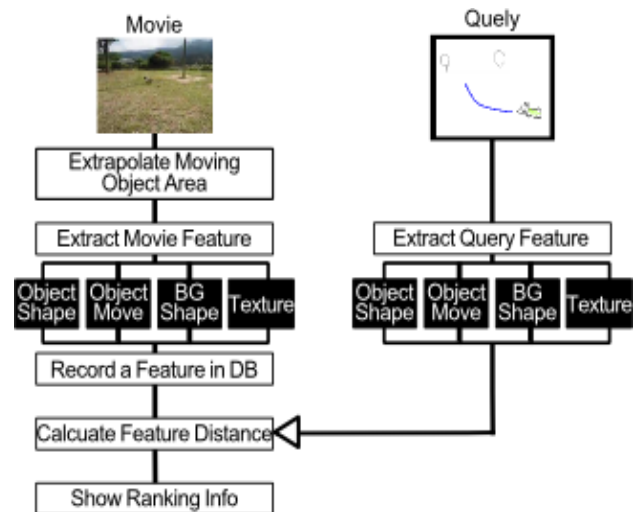


Figure 3. System flow of the proposed system.

A. SYSTEM REQUIREMENTS

In the proposed video retrieval system, target videos must meet the following requirements.

- Feature only one main subject
- Include a moving object
- Include camera work (pan, tilt)
- Not include editing (scene changes, caption insertions, etc.)

Conventional methods also have these same conditions.

B. SEARCH QUERY

Our system query is based on a hand-drawn sketch, an example of which is shown in Fig.4. Our system query is similar to Sekura's and differs in that we have added a texture input part. The user designates a purpose

(e.g., object shape, object movement, background shape, texture) to each sketch. Thus, needless checking drawing purpose and separate each method later.

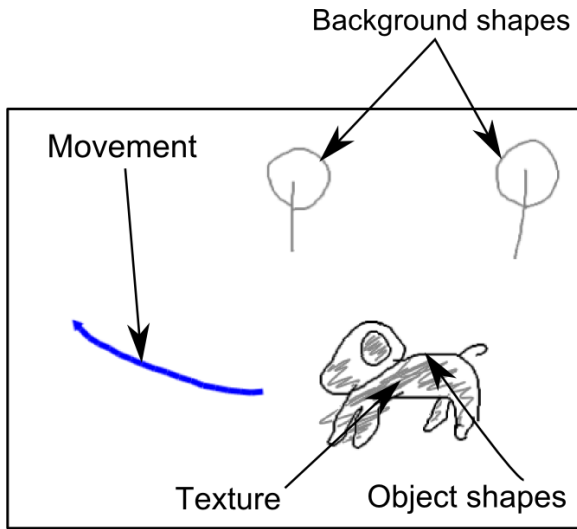


Figure. 4. Query example

C. TEXTURE FEATURE

Texture features (color and pattern) are related to the depicted texture information in a sketch. We focused our attention on color information from the texture depiction. Texture in a drawing is influenced by brightness, which we already know from Igarashi's study, so in our study, the texture features are also related to brightness. We used texture features and moving objects to determine areas with average brightness. For texture features used in a query, we use texture density for the object shape description. We also use extraction pattern information on a trial basis.

Next, we describe the texture feature extraction method. We extract texture features from several independent videos and average the values. This is because the brightness of a moving object tends not to change much in our target video. Now, we describe how the features are extracted in each frame. First, we calculate the pixel brightness of the moving object area using the same method as the conventional method. Next, we calculate the average brightness of pixels in the moving object area. Finally, we divide the average brightness by the maximum brightness value (normalizing). This value is then the texture feature of the target frame.

We extract the texture density from the sketch. First, the object shape of the drawn area is extracted from the sketch. Next, we check the texture description of each pixel in the extraction area, and finally, we calculate the texture density by dividing the texture description pixel number by the extraction area pixel number. This density percentage is the texture feature of the sketch.

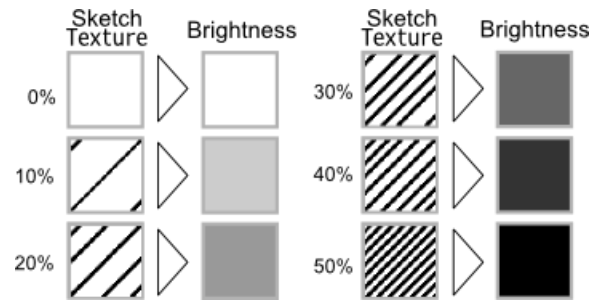


Figure. 5. Feature Correspondence

Figure.5 shows examples of sketch feature correspondence between a candidate video and a sketch. The higher the texture density, the lower the correspondence brightness. In a hand-drawn sketch, fully filled texture is very rare. Therefore, if the texture density is over half full, the correspondence brightness is the lowest.

We use the following expression to calculate the texture feature distance.

$$D_T = |D_{TV} - D_{TS}| \quad (1)$$

Where D_{TV} is a texture feature from a video and D_{TS} is a texture feature from a sketch. In this way we can calculate the absolute value of the differences between the movie feature and the sketch feature.

Next, we describe the pattern information. Extracted pattern information is performed on a trial basis because pattern drawing differs greatly among different users. We use the number of pattern information, which is one of the simplest methods.

Next, we describe the extraction method. From the video, we use the count from the low-brightness region. (The region is recognized by a visual evaluation.) From the sketch, we use the number of lines used to draw the texture because ordinarily in a rough sketch one pattern is drawn by one line.

We use the following expression to calculate the number of pattern difference D_p .

$$D_p = \min_{1 \leq n \leq N} |D_{pV_n} - D_{pS}| \quad (2)$$

where D_{pV} is the number of pattern information from video n frame, D_{pS} is the number of pattern information from the sketch, and N is the number of the final video frame. We calculate the difference of each movie number and sketch number. The highest least-absolute value in these calculated differences was D_p .

D. FINAL DISTANCE

The difference between video and sketch feature distance is calculated by the following expression.

$$D = C_1 D_V + C_2 D_F + C_3 D_B + C_4 D_T \quad (3)$$

where D_V is the movement feature distance, D_F is the object shape feature distance, D_B is the background shape feature distance, and D_T is the texture feature distance.

D_V , D_F , and D_B are computed using the same method as Sekura.

$C_{1,2,3,4}$ is variable for weight, and $C_1 = 1.0, C_{2,3,4} = 0.5$ in our work. These values are based on the existing system, and therefore the movement feature is maximal one. Our proposed feature has the same weight as the Sekura[1] proposed.

We use the following expression to add the number of pattern differences:

$$D = C_1 D_V + C_2 D_F + C_3 D_B + C_4 D_T + C_5 D_P \quad (4)$$

where D_V , D_F , D_B , D_T , and $C_{1,2,3,4}$ are the same as above.

D_P is the number of pattern difference.

C_5 is 10.0 in our work because the texture pattern drawing is an intentional act and we felt that texture pattern information is therefore the most important feature.

IV. EXPERIMENT & DISCUSSION

In this section, we describe our experiment using the texture feature. In the experiment, we compared the proposed system with the conventional one in terms of how well it could retrieve video both with and without the texture feature.

A. EXPERIMENT METHOD

Here, we describe the experiment method.

First, we set the retrieval goal video from among various retrieval target videos.

Next, we made our sketch query after watching the goal video.

Finally, we attempted to retrieve the movie with both systems using the sketch query.

When using the existing system, we removed texture information from the query.

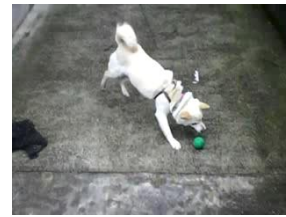
We conducted ten trials and then compared the retrieval results of the two systems.

B. TARGET MOVIES

Here, we describe the target movies used in the experiment. We searched YouTube with key words

"dog, ball, playing". From the results, we selected videos in which the color of the moving object (the dog) was clear and that fulfilled all the prerequisites listed in Section 3.1. We did not select any dappled dog videos or any in which dogs were dressed up because the texture feature is designed to distinguish colors in moving objects; it cannot function with dappling or with dog clothes.

We selected ten videos to use as experiment Set 1. The intent with this set was to verify the texture feature. Set 1 includes four white dog videos (Videos A, B, C, and D), four black dog movies (Videos E, F, G, and H), and 2 brown dog movies (Videos I and J). An example from each is shown in Fig.6. The dog color was determined by visual confirmation.



(a) White (Video D)



(b) Black (Video F)



(c) Brown (Video J)

Figure 6. Search target movie samples

We extracted texture feature from these selected videos. We then extracted texture features from the selected videos. The extraction results are shown in Table.1.

Table 1. Extracted Texture Feature

Moving Object Color	Texture Feature(lowest - highest)
White	0.5383 - 0.6378
Black	0.1925 - 0.3003
Blown	0.2793 - 0.4560

Generally there were no problems with extracting the features. However, the accuracy of the texture feature in Video I was lower than in the black ones.

We prepared another video set to verify the number of pattern differences. For Set 2, we added three videos to Set 1, all of which had clear low-brightness areas.

C. RESULT & DISCUSSION

In this section, we discuss the experimental results of

the proposed system.

Let us look at the results of Set 1 first. With Set 1, we used expression(3). Our texture feature can discriminate the color of moving objects in video, as shown in Table.1 , which suggests that texture features have potential for use in video retrieval. However, in the experiment, both strong and poor results were seen. We discuss the reason behind this with an example from each result.

First, we discuss the case in which the proposed system target video ranking was higher than that of the conventional system. It was especially higher in videos A and E. Table.2 shows how the target video rank changed.

Table.2 Example of highly ranked target videos

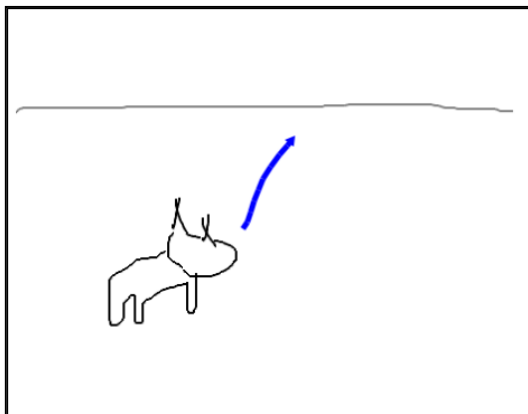
Target Video	Conventional method rank	Proposed method rank
Video A	4	2
Video E	3	1

For each example, there is not much difference in terms of feature distance with the conventional method. This is because it used a similar scene query as the movies in Fig.7. Therefore, there were more texture feature differences with the proposed system.

Second, we discuss the case in which the proposed system target video ranking was lower than the conventional system. Table.3 shows how the target video rank changed.

Table.3 Example of lowly ranked target videos.

Target Video	Conventional method rank	Proposed method rank
Video H	2	5
Video I	1	5



(a) Query



(b) Video A (Target Video)

(c) Video H (Similar Video)

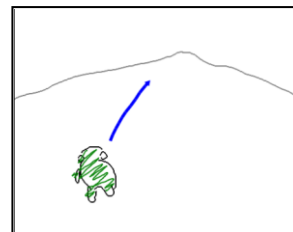
Figure 7. Video A query and similar scene

Figure.8 shows one scene of a target video that was retrieved using Video H and Video I queries. In each example, the texture information was input in query, but the density was not high (about 20\%). Therefore, the texture feature value is about 0.6, and this value is white rather than black or brown. This means that when the moving object color is black and brown, the target videos have a lower rank, and when the moving object is white, the video has a higher rank.

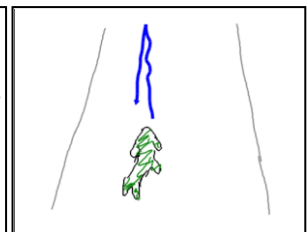


(a) Video H scene

(b) Video I scene



(c) Video H query



(d) Video I query

Figure 8. Video H and I query and one video scene?

In this case, inputting the appropriate texture feature is too difficult for our retrieval system. We need to develop a more appropriate texture feature input method to improve the retrieval result.

Sometimes the target video rank changed very little. These cases tended to have one of two patterns. First, the target video could be sufficiently distinguished only with the conventional system because our texture feature was too small to much change the result. Second, the scene was the same as that of the moving object color in another video, which could not be distinguished by the proposed system.

Throughout the experiment from these things, our texture feature is useful feature for distinguish video, it can't existing system. Second , we discuss experimental Set 2, which used expression (4). As seen in Table.4, the number of pattern differences can be useful in terms of

improving the retrieval accuracy. However, these results were from cases in which the pattern number was inputted correctly. If we use the number of the pattern in an actual environment, we need to assume that the user input will be wrong in some cases.

Table.4 Number of pattern include evaluation

Rank movement (Compare to conventional method)	Number of movies
Higher rank	7
No change	6
Lower rank	0

V. CONCLUSION

We proposed a video retrieval system that uses texture feature to improve the retrieval precision. We focused on texture information (e.g., rough fill, pattern) in a hand-drawn sketch. We proposed texture features based on texture information in the sketch and on video brightness. Results showed that using texture features in our proposed system has the potential to improve video retrieval. However, we had a problem texture information input. In addition, in this study, when extracting texture features from video, we checked only the brightness of the moving object area. If what goes on, texture features cannot be extracted accurately in cases in which all areas of a video are either high- or low-brightness video. Accordingly, we need to consider background brightness when extracting texture features.

For future work, we intend to add pattern information to texture information to assist with retrieval. We found that pattern information is useful information for retrieval. However, pattern information weight in evaluation expression is unclear, and we need more experimental data.

ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Scientific Research (C) (22500111).

REFERENCES

- [1] Akihiro Sekura and Masashi Toda: "Video Retrieval System Using Handwriting Sketch", Proc. of the IAPR Conference on Machine Vision Applications (MVA2009), pp.283-286, 2009.05.
- [2] Yuichi Igarashi and Masashi Toda: "A study on Image Search by Handwriting Sketch which Considered a Texture", Proc. of 2011 IEICE Hokkaido Internet Symposium, pp.49-53, 2011.
- [3] J.P.Collomosse, G.McNeill, L.Watts, "Free-hand sketch grouping for video retrieval", Proc. of IEEE International Conference on Pattern Recognition (ICPR 2008), pp.1-4, 2008.
- [4] M.Thorne, D.Burke, M.Panne, "Motion Doodles A Sketch-based Interface for Character Animation", SIGGRAPH 2004 courses, vol. 23, vol. 3, pp.175-182, 2004.
- [5] YouTube: <http://www.youtube.com/>
- [6] O. Gosuke, N. Yasutake, M. Keita and S. Yoshifumi: "Edge-based Image Retrieval Using a Rough Sketch", The Journal of the Institute of Image Information and Television Engineers, Vol.56, No.4, pp.653-658, 2002 (in Japanese).
- [7] M. Matuzaki, M. Kasimura and S. Ozawa: "Image Retrieval for Pictures Based on Feature Graph Using Simplified Image", The transaction of the Institute of Electronics, Information and Communication Engineers. D-II, vol.J87-D-II, No.2, pp.521-533, 2004 (in Japanese).
- [8] H. Takahiro, I. Atsushi and O. Rikio: "Retrieval of Landscape Images Using a User-drawn Sketch and Semantic Icons", Technical report of IEICE. PRMU, vol.104, No.573, pp.19-23, 2005.
- [9] M. Dan, A. Shozo and M. Masashi: "An appearancebased video scene retrieval based on synthesized video query by target points indication and the line history image matching". Technical report of IEICE. PRMU, vol.106, No.606, pp.67-72, 2007 (in Japanese).
- [10] F. Yasuhiko, M. Takamichi, I. Yasuhiro, K. Aki, Y. Katunori and S. Yoshihiro: "Video Retrieval System with Query Generation Assistance Using Relevance Feedback", IEICE technical report. Image engineering, vol.105, No.610, pp.95-100, 2006 (in Japanese).
- [11] S. Kenichirou, N. Masaomi and S. Hitoshi: "Moving Object Detection from Video Sequence

Using Local Motion Direction Histogram Features", IEICE technical report. Neurocomputing, vol.102, No.382, pp.47-52, 2002 (in Japanese).

Hiroki Kobayashi, was received B.S. in Systems Information Science from Future University Hakodate. His research interest in schooldays is Computer Vision.

Masashi Toda, was born in Hamamatsu, Shizuoka Pref., Japan in 1969. He received B.S. from the University of Tokyo, Japan in 1993, and M.S. and Ph.D. in electro-informatics engineering from Hokkaido University, Japan in 1995 and 1998. From 1998 to 2001, he was a researcher in IS Labo., SECOM Co., Ltd., Japan. From 2001 to 2005, he was an assistant professor in School of Systems Information Science, Future University Hakodate, Japan. From 2005 to 2012, he was an associate professor in Future University Hakodate, Japan. Since 2012, he has been a professor in Center for Multimedia and Information Technologies, Kumamoto University, Japan. His main research interest is in image processing technology. He is also interested in human interface technology, wearable computing, ubiquitous computing, information retrieval technology, and educational information system. He is a member of the Information Processing Society of Japan.