

# Computational Metrics Can Provide Quantitative Values to Characterize Arthroscopic Field of View



Ryan H. Barnes, M.D., M. Leslie Golden, M.D., David Borland, Ph.D., Reed Heckert, M.D., Meghan Richardson, M.D., R. Alexander Creighton, M.D., Jeffrey T. Spang, M.D., and Ganesh V. Kamath, M.D.

**Purpose:** The purpose of this study was to determine the inter-rater reliability of arthroscopic video quality, determine correlation between surgeon rating and computational image metrics, and facilitate a quantitative methodology for assessing video quality. **Methods:** Five orthopaedic surgeons reviewed 60 clips from deidentified arthroscopic shoulder videos and rated each on a four-point Likert scale from poor to excellent view. The videos were randomized, and the process was completed a total of three times. Each user rating was averaged to provide a user rating per clip. Each video frame was processed to calculate brightness, local contrast, redness (used to represent bleeding), and image entropy. Each metric was then averaged over each frame per video clip, providing four image quality metrics per clip. **Results:** Inter-rater reliability for grading video quality had an intraclass correlation of .974. Improved image quality rating was positively correlated with increased entropy (.8142;  $P < .001$ ), contrast (.8013;  $P < .001$ ), and brightness (.6120;  $P < .001$ ), and negatively correlated with redness (-.8626;  $P < .001$ ). A multiple linear regression model was calculated with the image metrics used as predictors for the image quality ranking, with an R-squared value of .775 and root mean square error of .42. **Conclusions:** Our study demonstrates strong inter-rater reliability between surgeons when describing image quality and strong correlations between image quality and the computed image metrics. A model based on these metrics enables automatic quantification of image quality. **Clinical Relevance:** Video quality during arthroscopic cases can impact the ease and duration of the case which could contribute to swelling and complication risk. This pilot study provides a quantitative method to assess video quality. Future works can objectively determine factors that affect visualization during arthroscopy and identify options for improvement.

## Introduction

Arthroscopic surgery is growing in frequency and comprises some of the most commonly performed orthopaedic surgeries.<sup>1</sup> Arthroscopy enables minimally invasive techniques with improved intra-articular visualization and potentially a quicker recovery when compared to open surgery.<sup>2</sup> Technology and

instrumentation associated with arthroscopic orthopaedic surgery have drastically changed and progressed over the last few decades, with improvements in video and monitor quality to 4K resolution, decreased scope size and smaller instruments, closed loop pumps, and improvements in image enhancement, lens, and fiber-optics. The basic setup typically involves an arthroscopic tower, which houses various power boxes for motorized instruments and irrigation pumps, viewing monitors, a light source with a light cable, and an arthroscope and camera (Fig 1).<sup>3</sup> Arthroscopes are telescopic cameras that contain magnifying lens systems within a tube. Video resolution depends on multiple factors, including the light source and transmission, lens, and monitor quality.

Separate from arthroscopy system characteristics, such as image resolution, the quality of the surgical visualization can change drastically on the basis of bleeding within the operative field. Bleeding can be controlled via pump pressure and through other external factors such as hypotensive anesthesia and the use of diluted epinephrine within arthroscopy

University of North Carolina at Chapel Hill, Department of Orthopaedics, Chapel Hill, North Carolina, U.S.A. (R.H.B., M. L.G., R.H., M.R., R.A.C., J.T.S., G.V.K.); and University of North Carolina at Chapel Hill, Renaissance Computing Institute (RENCI), Chapel Hill, North Carolina, U.S.A. (D.B.).

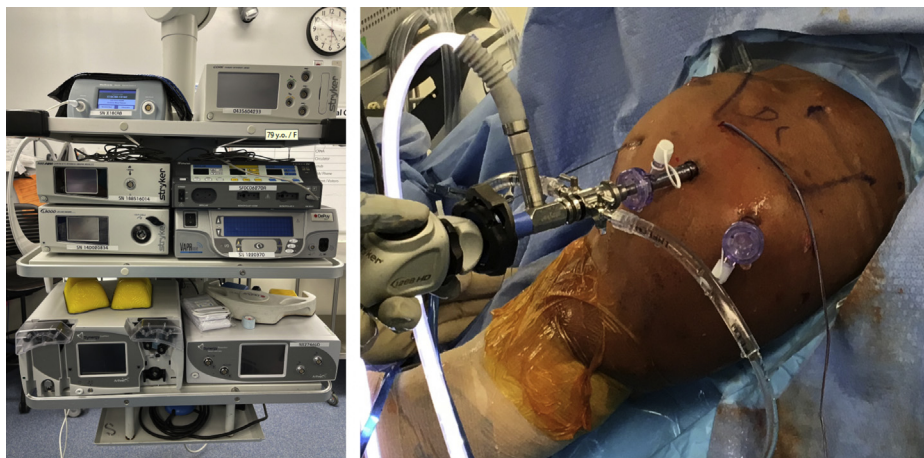
The authors report the following potential conflicts of interest or sources of funding: A.C. reports fellowship support and being a paid consultant for Arthrex; he has received fellowship support from Smith & Nephew and Breg. Full ICMJE author disclosure forms are available for this article online, as supplementary material.

Received June 8, 2021; accepted October 24, 2021.

Address correspondence to Ryan H. Barnes, M.D., 130 Mason Farm Rd., Chapel Hill, North Carolina, 27514, U.S.A. E-mail: rhbarnes44@gmail.com

© 2021 THE AUTHORS. Published by Elsevier Inc. on behalf of the Arthroscopy Association of North America. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). 2666-061X/21829

<https://doi.org/10.1016/j.asmr.2021.10.017>



**Fig 1.** Standard arthroscopic setup at our institution.

fluid.<sup>4,5,6,7,8</sup> Better arthroscopic field of view can enable improved differentiation between structures within the field of view, thus improving efficiency and the ease associated with the procedure. On the other hand, reduced image quality can impede the procedure, increase operative time, and possibly result in complications associated with the surgery, such as increased blood loss and postoperative swelling.

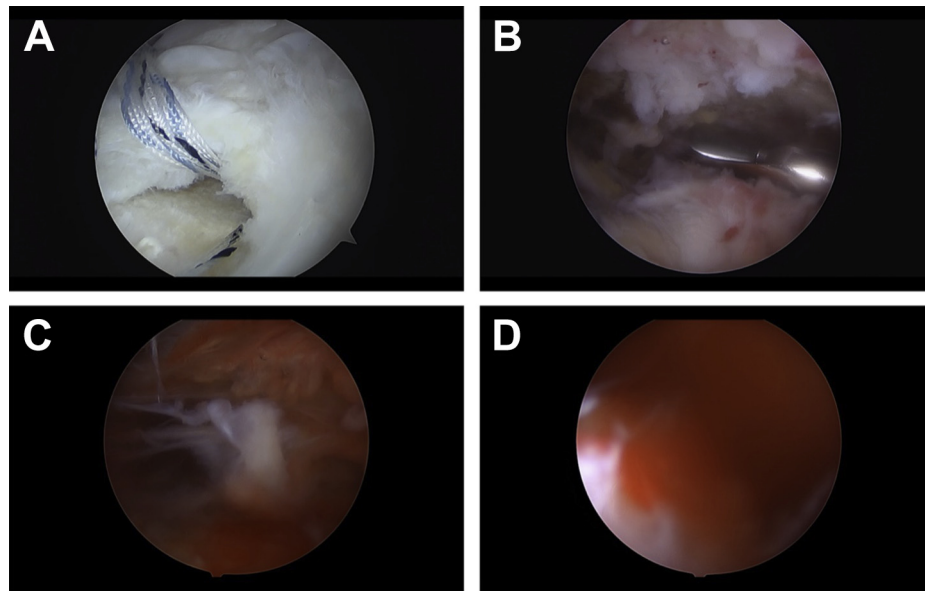
Typically, the quality of arthroscopic video has been qualitatively described by the surgeon. In this work, we investigate the use of general image quality metrics (brightness, local contrast, and image entropy), along with a metric (redness) to identify a specific factor (bleeding), impacting arthroscopic video quality. There is limited literature available regarding computer quantitative descriptions of arthroscopic videos. The purpose of this study was to determine the inter-rater reliability of arthroscopic video quality, determine correlations between surgeon rating and computational image metrics, and facilitate a quantitative methodology for assessing video quality. Our hypotheses were that there would be a strong inter-rater reliability among surgeons, the three general image quality metrics (brightness, local contrast, and image entropy) would positively correlate with user ratings of video quality, and the redness metric would negatively correlate with user ratings of video quality.

## Methods

Institutional Review Board approval (IRB; no. 20-2913) was obtained for this study. A total of 60 deidentified arthroscopic video clips were selected. These videos ranged in duration from 7 seconds to 16 seconds (average length of 11.28 seconds) and were taken from a total of four shoulder procedures. All videos were obtained from shoulder arthroscopy cases. Knee arthroscopy cases were not included in our study, as a tourniquet is commonly used to decrease bleeding. The specific videos were selected, as we felt that they best represented different

levels of bleeding despite using a similar equipment setup. All cases were performed in the beach chair position and used the same arthroscopic equipment, which captured and recorded videos with no difference in setup video quality: using a Stryker (Kalamazoo, MI) 1588 High Definition 4 mm × 30-degree arthroscope with a Stryker “T” trocar handpiece with one-way inflow and one-way outflow design and Arthrex (Naples, FL) dual-flow pump. The surgeries were performed by a single, fellowship-trained sports medicine orthopaedic surgeon (G.V.K.). Pump pressure was controlled for all surgeries, with the pump pressure set at 50 mmHg, with lavages as needed during the case, which increases the pressure up to 50% for 120 seconds. Epinephrine was used in the first two fluid bags. Blood pressure management was deferred to anesthesia with a preference for hypotensive anesthesia when able to be tolerated. The videos covered a range of visualization quality due to extrinsic factors, such as bleeding. Five orthopaedic surgeons (three attending sports medicine orthopaedic surgeons and two sports medicine orthopaedic surgery fellows) reviewed each of the 60 deidentified arthroscopic video clips. Each clip was rated on a four-point scale from poor view to excellent view. For the purposes of data analysis, a previously used grading scale was used with a poor view given a value of 1, and an excellent view given a value of 4 (Fig 2).<sup>9</sup> The videos were randomized, and the process of viewing and grading was completed a total of three times per reviewer with at least 2 months between each assessment. For each video, the 15 ratings (5 raters, 3 ratings each) were averaged to calculate a single rating.

Brightness was selected to capture the general illumination level during the procedure, which could be affected by factors, such as bleeding. Local contrast was selected to capture the amount of fine detail in the scene. The human visual system is most sensitive to changes in intensity when processing small-scale details. Redness was selected to capture a specific factor thought to impact arthroscopic visual quality and bleeding.



**Fig 2.** Representative screenshots from videos that received unanimous grading for visualization quality by all 5 participants. (A) Unanimous score of 4: excellent view—no limitation of view, procedure unimpeded. This screen shot is taken from the beach chair position using the posterolateral viewing portal, looking at the rotator cuff during a rotator cuff repair. (B) Unanimous score of 3: good view—slightly limited, procedure unimpeded. This screen shot is taken from the beach chair position using posterolateral viewing portal, looking at acromion during acromioplasty. (C) Unanimous score of 2: fair view—limited, procedure impeded slightly. This screen shot is taken from the beach chair position using the posterolateral viewing portal. (D) Unanimous score of 1: poor view—limited, procedure impeded markedly. This screen shot is taken from the beach chair position using the posterolateral viewing portal, looking at the subacromial space during a rotator cuff repair.

### Statistical Analysis

Statistical analysis was performed with intraclass correlation (ICC) to calculate the inter-rater reliability (IRR), as well as *t*-test with an  $\alpha$  of .05, power of .8, and a moderate effect size of

.3.<sup>10</sup> Power analysis was performed to determine the number of clips to provide appropriate power. Video processing and analysis were conducted using the MatLab computing platform. Each video was initially processed using a Hough transform to compute a mask of the circular view area, indicating valid pixels for further processing (Fig 3).<sup>11</sup> The four image quality metrics were then computed for each frame, as follows: brightness: mean pixel grayscale intensity; local contrast: computed per pixel as the standard deviation of the grayscale pixel intensities in the  $3 \times 3$  neighborhood centered on the pixel, and then averaged per frame; image entropy: computed on the grayscale pixel intensities as  $-\text{sum}[p \cdot \log_2(p)]$ ; and redness: the redness per pixel is computed from the (r, g, b) components as  $\max[0, r - (g + b) / 2]$ , and then averaged per frame. Image entropy is a statistical measure of randomness that is often interpreted as the degree of information content in an image.<sup>12,13</sup> Figure 4 demonstrates the difference between the same images, but with differing entropy. Each metric was then averaged across each frame from a video clip to calculate the quality metrics for that video. Correlations (Pearson's *r*) were

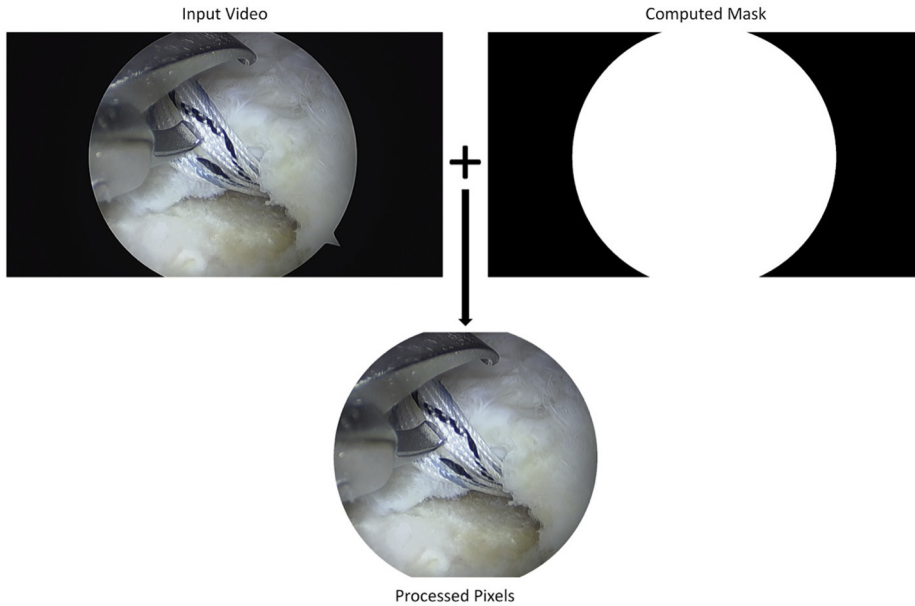
calculated for each rating/image metric pair. Subsequently, a multiple linear regression model was computed using the four image quality metrics to predict the user rating.

### Results

Scatterplots of the videos were created for each user rating/image metric pair, and the corresponding correlations calculated. Improved image quality rating was positively correlated with increased brightness, contrast, and entropy, and negatively correlated with redness (Fig 5, Table 1, and Table 2). All correlations represented statistically significant values. The inter-rater reliability (IRR) was calculated using intraclass correlation (ICC), with a value of .974 (95% confidence interval:  $.963 < \text{ICC} < .982$ ). The model achieved an R-squared value of .83, with a root mean squared error (RMSE) of .368. K-fold cross validation was also performed, achieving an RMSE of .4066 (Fig 6).

### Discussion

Our study demonstrates strong inter-rater reliability between surgeons when describing image quality, as well as strong correlations between image quality and image metrics, confirming our hypothesis. A more quantitative, objective assessment of arthroscopic video quality may be practically impactful as a way to confirm



**Fig 3.** Illustration of masking process used to identify valid pixels for processing in each video frame.

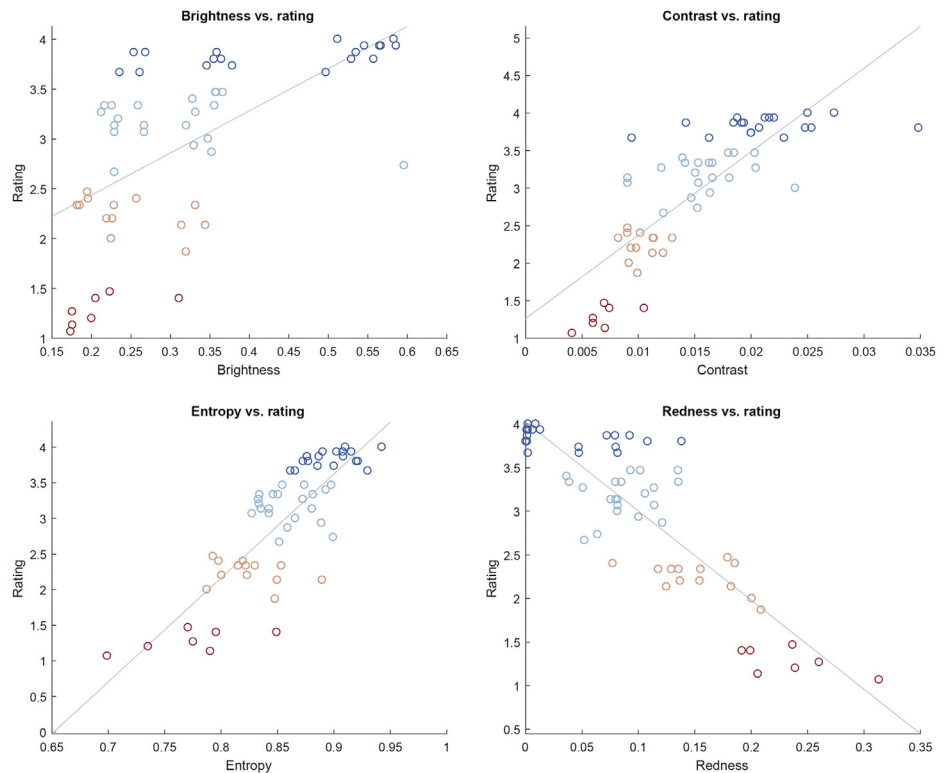
the effectiveness of commonly used interventions that are believed to improve arthroscopic field of view. This study builds upon computational literature but applies it to assess a new field in orthopaedic arthroscopic video quality. This computational method to assess video quality provides a new way to describe image metrics to arthroscopic visualization in what has only been described qualitatively. The results indicate that our model is able to explain 83% of the variance in the

data, with an error less than half of a point on the rating scale. K-fold cross validation was also performed, achieving an RMSE of .4066, indicating that the model should generalize to new data.

Optimizing visualization of the surgical field is paramount to successful arthroscopy and depends on many factors, including pathology addressed during the case, surgeon experience or technique, and extrinsic factors. Multiple techniques have been used in an attempt to

**Fig 4.** Two images to demonstrate image entropy. The first image has entropy of 5.5128. The second is a posterized version of the first, with a reduced the number of grayscale values in the image, which has entropy of 1.5023.





**Fig 5.** Scatterplots for each image metric against average user rating, with least squares line of best fit. Each circle represents a video and is colored by average user rating value.

improve visualization, including tourniquet use, thermal electrocautery, digital pressure over portals, pump pressure settings, and arthroscopic fluid adjuncts.<sup>14,15,16</sup> Despite the benefit these various techniques provide with improvement in arthroscopic visualization, they are not without their potential side effects or complications and are well documented, including increased swelling and chondrolysis.<sup>17</sup> Other attempts to improve visualization have been other extrinsic factors, including hypotensive anesthesia and epinephrine in fluid bags, which also have their share of potential complications.<sup>18,19</sup> Previous studies from Avery et al., Jensen et al., van Moortfoot et al., and Kuo et al. have demonstrated that external options, such as hypotensive anesthesia and epinephrine in fluid bags to control bleeding and, in turn, visualization can be beneficial.<sup>4,5,6,7</sup> However, these studies are based on qualitative descriptions of visual quality based on either visual analog scales or numeric rating scales. Our study can be

applied to these studies to provide quantitative support of their qualitative findings through our model. This study may impact future studies by providing a foundational framework to quantitatively describe and confirm the effect of interventions commonly used to improve arthroscopic field of view.

**Limitations**

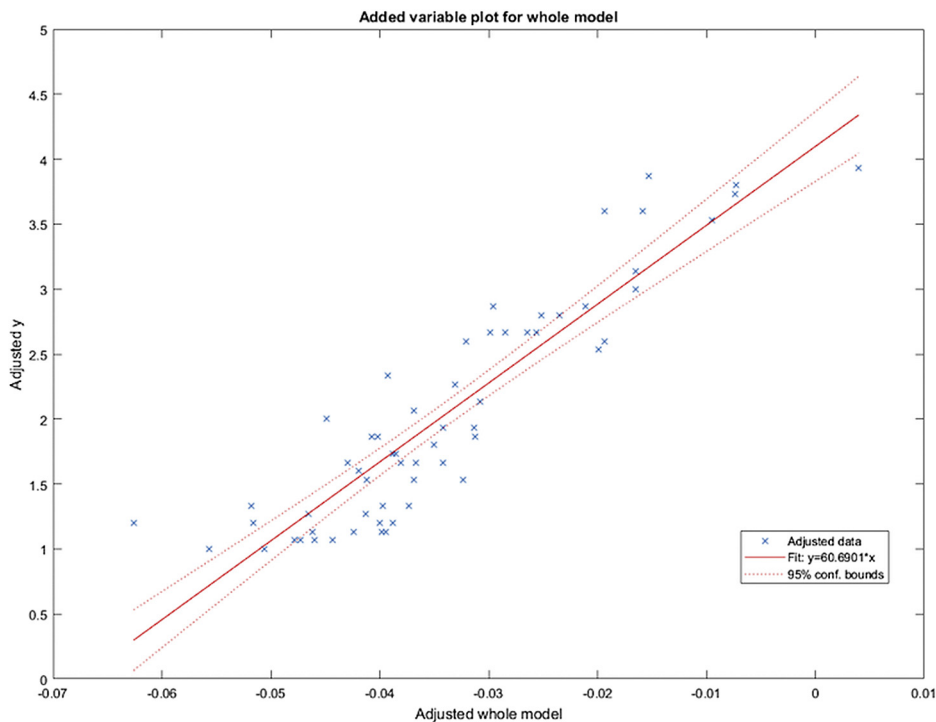
There are several limitations to this study. First, there were only a total of 5 participants. More participants could change the inter-rater ratings. Second, although the results of the K-fold cross-validation indicated that the multiple linear regression model should generalize to new data, it is unclear whether this extends to different arthroscopic imaging systems, which may have different image resolution and optics. The image entropy metric in particular may be sensitive to such intrinsic factors, so future work should include examining the generalizability of these metrics across imaging systems. Also, only a single video system was used

**Table 1.** Correlation Value (Pearson’s *r*) and *P* Value for Each Image Metric Pair

Image Metrics Correlations	Correlation Value ( <i>r</i> )	<i>P</i> Value
Redness versus brightness	-.6027	<.001
Redness versus contrast	-.6518	<.001
Redness versus entropy	-.7380	<.001
Brightness versus contrast	.7049	<.001
Brightness versus entropy	.7469	<.001
Contrast versus entropy	.7522	<.001

**Table 2.** Correlation Value (Pearson’s *r*) and *P* Value for Average User Rating and Each Image Metric

Image Metrics Correlations	Correlation Value ( <i>r</i> )	<i>P</i> Value
Redness versus rating	-.8626	<.001
Brightness versus rating	.6120	<.001
Contrast versus rating	.8013	<.001
Entropy versus rating	.8142	<.001



**Fig 6.** Added variable plot of whole linear regression model, illustrating that the model is significant.

during all surgeries, and it is challenging to control for other variables consistently between cases, such as blood pressure and pump pressure. Another limitation is that redness could be impacted by factors other than bleeding, including pathology being addressed and location of the scope (subacromial space vs visualizing the central portions of the glenohumeral joint), which could negatively impact subjective assessment of image quality. Finally, although we have established quantitative measurements for these variables, we have not defined benchmarks to apply to show how these measurements and variables are clinically relevant, such as how a particular amount of entropy adds to the duration of time for an arthroscopic procedure, or makes it more challenging.

## Conclusions

Our study demonstrates strong inter-rater reliability between surgeons when describing image quality and strong correlations between image quality and the computed image metrics. A model based on these metrics enables automatic quantification of image quality. We believe that our results are generalizable to all joints. With the use of these metrics, future studies can move to objective values for video quality. This pilot study represents a usable, quantitative tool for assessing the effect of extrinsic factors on arthroscopic visualization (i.e., pump pressure, hypotensive anesthesia, and epinephrine in fluid bags) in further works, which can further expand and confirm conclusions

from previous visualization studies that relied solely on visual analog scales or numeric rating scales.

## References

1. Kim S, Bosque J, Meehan JP, Jamali A, Marder R. Increase in outpatient knee arthroscopy in the United States: A comparison of National Surveys of Ambulatory Surgery, 1996 and 2006. *J Bone Joint Surg* 2011;93:35-41.
2. Paxton ES, Backus J, Keener J, Brophy RH. Shoulder arthroscopy: Basic principles of positioning, anesthesia, and portal anatomy. *J Am Acad Orthopaed Surg* 2013;21:332-342.
3. Randelli PS, Cucchi D. Surgery set-up, instrumentations and electronic equipment. In: Randelli P, Dejour D, van Dijk CN, Denti M, Seil R, eds. *Arthroscopy basic to advanced*. New York: Springer, 2016;13-19.
4. Van Montfoort DO, van Kampen PM, Huijsmans PE. Improvement of clarity of visual field and shortening of total operation time. A randomized controlled trial. *Arthroscopy* 2016;32:436-444.
5. Avery DM III, Gibson BW, Carolan GF. Surgeon-rated visualization in shoulder arthroscopy: A randomized blinded controlled trial comparing irrigation fluid with and without epinephrine. *Arthroscopy* 2015;31:12-18.
6. Jensen KH, Werther K, Stryger V, Schultz K, Falkenberg B. Arthroscopic shoulder surgery with epinephrine saline irrigation. *Arthroscopy* 2001;17:578-581.
7. Kuo LT, Chen CL, Yu PA, Hsu WH, Chi CC, Yoo JC. Epinephrine in irrigation fluid for visual clarity in arthroscopic shoulder surgery: A systematic review and meta-analysis. *Int Orthopaed* 2018;42:2881-2889.

8. Ware C. *Information visualization: perception for design*. Burlington, MA: Morgan Kaufmann, 2019.
9. Johnson DS, Stewart H, Hirst P, Harper NJN. Is tourniquet use necessary for knee arthroscopy? *Arthroscopy* 2000;16: 648-651.
10. Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quant Methods Psychol* 2012;8:23-34.
11. Duda RO, Hart PE. Use of the Hough transformation to detect lines and curves in pictures. *Comm ACM* 1972;15:11-15.
12. Shannon CE. A mathematical theory of communication. *The Bell Syst Tech J* 1948;27:379-423.
13. Gonzalez RC, Woods RE, Eddins SL. *Digital Image Processing Using MATLAB, Chapter 11*. New Jersey: Prentice Hall, 2003.
14. Burkhart SS, Danaceau SM, Athanasiou KA. Turbulence control as a factor in improving visualization during subacromial shoulder arthroscopy. *Arthroscopy* 2001;17: 209-212.
15. Tuijthof GJM, Dusee L, Herder JL, van Dijk CN, Pisteccky PV. Behavior of arthroscopic irrigation systems. *Knee Surg Sports Traumatol Arthrosc* 2005;13:238-246.
16. Furia JP, Zambetti GJ Jr. An injection technique to create a bloodless field in arthroscopically assisted anterior cruciate ligament reconstruction. *Am J Sports Med* 1992;20: 406-409.
17. Baker JF, Mulhall KJ. Local anaesthetics and chondrotoxicity: What is the evidence? *Knee Surg Sports Traumatol Arthrosc* 2012;20:2294-2301.
18. Karaoglu S, Dogru K, Kabak S, Inan M, Halici M. Effects of epinephrine in local anesthetic mixtures on hemodynamics and view quality during knee arthroscopy. *Knee Surg Sports Traumatol Arthrosc* 2002;10:226-228.
19. Mazzocca AD, Meneghini RM, Chhablani R, Badrinath SK, Cole BJ, Bush-Joseph CA. Epinephrine-induced pulmonary edema during arthroscopic knee surgery. A case report. *J Bone Joint Surg Am* 2003;85: 913-915.